Lung-Hsiang Wong · Yugo Hayashi ·
Cesar A. Collazos · Claudio Alvarez ·
Gustavo Zurita · Nelson Baloian (Eds.)

LNCS 13632

Collaboration Technologies and Social Computing

28th International Conference, CollabTech 2022
Santiago, Chile, November 8–11, 2022
Proceedings

Springer

Lecture Notes in Computer Science 13632

More information about this series at https://link.springer.com/bookseries/558

Lung-Hsiang Wong · Yugo Hayashi ·
Cesar A. Collazos · Claudio Alvarez ·
Gustavo Zurita · Nelson Baloian (Eds.)

Collaboration Technologies and Social Computing

28th International Conference, CollabTech 2022
Santiago, Chile, November 8–11, 2022
Proceedings

Springer

Editors

Lung-Hsiang Wong
Nanyang Technological University
Singapore, Singapore

Yugo Hayashi
Ritsumeikan University
Osaka, Japan

Cesar A. Collazos
Universidad del Cauca
Popayán, Colombia

Claudio Alvarez
Universidad de Los Andes
Santiago, Chile

Gustavo Zurita
University of Chile
Santiago, Chile

Nelson Baloian
University of Chile
Santiago, Chile

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-031-20217-9 ISBN 978-3-031-20218-6 (eBook)
https://doi.org/10.1007/978-3-031-20218-6

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

Preface

This volume contains the papers presented at the 28th International Conference on Collaboration Technologies and Social Computing (CollabTech 2022). The conference was held during November 8–11, 2022, in Santiago, Chile, in hybrid format.

This year we received 38 submissions, including 26 full papers and 11 work in progress (WIP) contributions. Each paper was carefully reviewed by three to five Program Committee members in a double-blind process. As a result, the committee decided to accept 18 full and 4 WIP papers. The accepted papers present relevant and interesting research works related to theory, models, design principles, methodologies, and case studies that contribute to a better understanding of the complex interrelations that exist at the intersection of collaboration and technology. The program also included two keynote presentations.

As editors, we would like to thank the authors of all submissions and the members of the Program Committee for their careful reviews. Our thanks also go to the Universidad de los Andes, Chile, and its Faculty of Engineering and Applied Sciences, for hosting and supporting the conference through the "Fondo de Ayuda a la Investigación" (FAI) program. We thank the Universidad de Chile, including the Department of Management Control and Information Systems and the Department of Computer Science, for their support. Our special thanks go to the Information Processing Society of Japan (IPSJ) for their sponsorship. Finally, we would like to acknowledge the effort of the organizers of the conference, as well as thank the Steering Committee for the opportunity to organize the conference and their help provided in the process.

September 2022

Lung-Hsiang Wong
Yugo Hayashi
Cesar A. Collazos
Claudio Álvarez
Gustavo Zurita
Nelson Baloian

Organization

Conference Co-chairs

Claudio Álvarez	Universidad de los Andes, Chile
Gustavo Zurita	Universidad de Chile, Chile
Nelson Baloian	Universidad de Chile, Chile

Program Co-chairs

Lung-Hsiang Wong	Nanyang Technological University, Singapore
Yugo Hayashi	Ritsumeikan University, Japan
César Collazos	Universidad del Cauca, Colombia

Steering Committee

Nelson Baloian	Universidad de Chile, Chile
Heinz Ulrich Hoppe	RIAS-Institute, Germany
Tomoo Inoue	University of Tsukuba, Japan
Minoru Kobayashi	Meiji University, Japan
Hideaki Kuzuoka	University of Tokyo, Japan
Hiroaki Ogata	Kyoto University, Japan

Program Committee

Carlos Alario-Hoyos	Universidad Carlos III de Madrid, Spain
Ishari Amarasinghe	Universitat Pompeu Fabra, Spain
Manuela Aparicio	Universidade NOVA de Lisboa, Portugal
Juan I. Asensio-Pérez	Universidad de Valladolid, Spain
Miguel L. Bote-Lorenzo	Universidad de Valladolid, Spain
Daniela Caballero	McMaster University, Canada
Juan Felipe Calderón	Universidad Andrés Bello, Chile
Maiga Chang	Athabasca University, Canada
Irene-Angelica Chounta	University of Duisburg-Essen, Germany
Hui-Chun Chu	Soochow University, Taiwan
Yannis Dimitriadis	University of Valladolid, Spain
Orlando Erazo	Universidad Técnica Estatal de Quevedo, Ecuador
Micaela Esteves	IPLeiria, Portugal
Jesus Favela	CICESE, Mexico

Benjamim Fonseca	UTAD/INESC TEC, Portugal
Kinya Fujita	Tokyo University of Agrculture and Technology, Japan
Cédric Grueau	Polytechnic Institute of Setúbal, Portugal
Naoko Hayashida	Fujitsu Laboratories Ltd., Japan
Atsuo Hazeyama	Tokyo Gakugei University, Japan
Davinia Hernandez-Leo	Universitat Pompeu Fabra, Spain
Heinz Ulrich Hoppe	RIAS-Institute, Germany
Satoshi Ichimura	Otsuma Women's University, Japan
Claudia-Lavinia Ignat	Inria, France
Indratmo Indratmo	MacEwan University, Canada
Tomoo Inoue	University of Tsukuba, Japan
Yutaka Ishii	Okayama Prefectural University, Japan
Kazuyuki Iso	NTT Corporation, Japan
Marc Jansen	University of Applied Sciences Ruhr West, Germany
Hyungseok Kim	Konkuk University, South Korea
Jongwon Kim	Gwangju Institute of Science and Technology, South Korea
Ralf Klamma	RWTH Aachen University, Germany
Arianit Kurti	Linnaeus University, Sweden
Wim Lamotte	Hasselt University, Belgium
Thomas Largillier	GREYC, France
Liang-Yi Li	National Taiwan Normal University, Taiwan
Chiupin Lin	National Tsing Hua University, Taiwan
Rene Lobo	Universitat Pompeu Fabra, Spain
Tun Lu	Fudan University, China
Wolfram Luther	University of Duisburg-Essen, Germany
Maíra Marques Samary	Boston College, USA
Sonia Mendoza Chapa	CINVESTAV, Mexico
Roc Meseguer	Universitat Politècnica de Catalunya, Spain
Marcelo Milrad	Linnaeus University, Sweden
Carmen Morgado	Universidade NOVA de Lisboa, Portugal
Satoshi Nakamura	Meiji University, Japan
Mamoun Nawahdah	Birzeit University, Palestine
Kazushi Nishimoto	Japan Advanced Institute of Science and Technology, Japan
Alexander Nolte	University of Tartu, Estonia
Sergio Ochoa	University of Chile, Chile
Masayuki Okamoto	Toyota Motor Corporation, Japan
Masaki Omata	University of Yamanashi, Japan

Gerald Oster	LORIA, Université de Lorraine, Inria Nancy-Grand Est, France
Elvira Popescu	University of Craiova, Romania
Matías Recabarren	Universidad de los Andes, Chile
Oscar Revelo Sánchez	Universidad de Nariño, Colombia
Armanda Rodrigues	Universidade NOVA de Lisboa, Portugal
Luis Rojas	Universidad del Bio-Bio, Chile
Nobuchika Sakata	Osaka University, Japan
Flavia Santoro	UERJ, Brazil
Rodrigo Santos	Universidad Nacional del Sur, Argentina
Claudio Sapateiro	Polytechnic Institute of Setúbal, Portugal
Hidekazu Shiozawa	Tamagawa University, Japan
Marcus Specht	Delft University of Technology, Netherlands
Stefan Trausan-Matu	University Politehnica of Bucharest, Romania
Andrea Vasquez	Universidad Técnica Federico Santa María, Chile
Cristina Villa-Torrano	Universidad de Valladolid, Spain
Aurora Vizcaíno	Universidad de Castilla - La Mancha, Spain
Bahtijar Vogel	Malmö University, Sweden
Hao-Chuan Wang	University of California, Davis, USA
Takashi Yoshino	Wakayama University, Japan
Takaya Yuizono	Japan Advanced Institute of Science and Technology, Japan
Alejandro Zunino	CONICET-ISISTAN, UNICEN, Argentina

Keynote Abstracts

Teachers as Designers and Orchestrators of Technology-Enhanced Collaborative Learning: Challenges and Solutions

Yannis Dimitriadis

Universidad de Valladolid, Spain

Design and orchestration of technology-enhanced collaborative learning can be very challenging for teachers or even instructional designers. This keynote presentation deals with design for effective and efficient collaborative learning, and how teachers as designers and orchestrators may be supported in complex ecosystems. We present the main challenges and solutions regarding conceptual and technological tools which may be developed, building on, and adapting to existing design knowledge. The talk will provide an overview of patterns, approaches, tools, and systems that should respect teachers' agency while taking advantage of complex computational approaches, typically based on artificial intelligence. We pay special attention to recent research on how learning analytics solutions may be designed and implemented using human-centered approaches, and how socially shared regulated learning may be better supported. Several illustrating examples will be shown drawing on the literature and the research work of the presented during the last 25 years. Some prominent pending issues will be posed that may guide future research in supporting teachers as designers and orchestrators.

25 Years of CSCWD: Review and Perspective

Weiming Shen

Huazhong University of Science and Technology (HUST), China

Industries and societies require new technologies to address increasingly complex design issues for products, structures, buildings, systems, processes, and services while meeting the high expectation of customers. Computer Supported Collaborative Work in Design (CSCWD) emerged in response to this requirement. With the rapid advancement of Internet and Web based technologies, CSCWD has been a very active R&D area in the past two and half decades. Recent developments of Cloud/Fog/Edge Computing, Internet of Things, Big Data, Blockchains, and Digital Twin technologies have brought new opportunities for CSCWD. This talk will presents a brief review of the 25-year history of CSCWD and discusses future research opportunities and challenges.

Contents

Human Aspects in Software Development: A Systematic Mapping Study

Luz Marcela Restrepo-Tamayo(✉) and Gloria Piedad Gasca-Hurtado

Universidad de Medellín, Carrera 87 No. 30-65, 50026 Medellín, Colombia
{lmrestrepo,gpgasca}@udemedellin.edu.co

Abstract. Software development is a process that requires a high level of human talent management to ensure its success. This makes it a topic of interest to the software industry and research. Considering this interest, it is evident the need to know the aspects that have been studied, how they have been measured, and what data analysis methods have been used. This paper presents an analysis of the human aspects associated with the software development process, identifying procedures and methods used to analyze data and its measurement. A systematic mapping with a sample of 99 studies identified by their relationship with the proposed topic was used as the research method. The main findings show that one of the most studied is personality. This aspect is related to the performance of software development teams and is a key variable for its conformation. Concerning the most used data source, we find the survey based on self-reporting. Finally, descriptive statistics is the most frequent method of analysis, which is performed prior to other methods such as correlation or regression analysis. The results suggest a wide spectrum of human aspects to be studied in Software Engineering, and interesting potential for analysis by identifying interesting methods other than self-reporting.

Keywords: Metrics · Human aspects · Software development · Systematic mapping study

1 Introduction

There is a growing interest in studying human aspects in Software Engineering (SE), mainly because software development is a people-centered process [1], therefore human aspects have become a fundamental part of their measurement programs [2]. Previous studies indicate that the people and tasks associated with team processes are key to the success and effectiveness of software development projects [3]. In addition, human aspects must be taken into account in the estimation of productivity [4].

Research related to the identification of relevant human aspects in software engineering has been carried out [5]. However, it has been identified that the measurement of these aspects is difficult and therefore represents a challenge for both industry and research [6]. Consequently, the objective of this study is to identify the human aspects that have been studied in the context of SE, analyzing measurement instruments and data processing methods.

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 1–22, 2022.
https://doi.org/10.1007/978-3-031-20218-6_1

Considering the objective mentioned above, a systematic mapping is performed with a sample of 99 studies published in a time window between 2010 and 2021. Scientific databases such as Science Direct, Springer, IEEE, and ACM were used. The descriptive analyses described in this work are related to the year of publication, authors' country of affiliation, type of study, human aspects treated as dependent variables, human aspects treated as independent variables, measurement instruments used, and data analyzing methods.

The findings of this mapping allow us to identify the human aspects most studied in SE. A company may focus investment efforts by identifying the instruments that have been used to measure human aspects of interest. A higher education institution or research center will be able to identify those human aspects that have not been worked on and allocate research resources to promote projects aimed at the expansion of knowledge in the area.

Another interesting finding is the identification of tools for capturing data and analyzing data methods related to the measurement of human aspects in SDT. This identification can guide the definition of innovative methodologies and proposals, with advantages over traditional ones, and that can be easily incorporated in academia and industry to improve teamwork.

This article is structured as follows. Section 2 is related to the importance of human aspects in software development. Section 3 presents the systematic mapping protocol followed in this article. Section 4 presents the overall results of the mapping and answers the research questions. Section 5 presents implications for practice and research, and the limitations of the study. The conclusions are in Sect. 6.

2 Human Aspects in Software Development

Human aspects have an important impact on SE [7]. A software product requires human intervention [8] and its development is an intellectually challenging activity that demands collaborative work [9].

Several studies indicate that SDT performance is affected by technical, non-technical (soft), organizational and environmental factors [10]. Non-technical factors include those related to people [11, 12].

Skills such as communication, emotional intelligence or leadership are required in jobs related to Information Technology [13]. However, organizations must understand the importance of these types of aspects, as well as diversify the skills of developers to enrich talent and contribute to the work of building software [14].

Since people-related factors have raised the interest of SE researchers, there is a set of 57 social and human factors that, according to a tertiary review, influence the productivity of SDT [5]. This set was adjusted to 13 factors by a conceptual analysis supported by psychology and software engineering [15] and corroborated through a survey-based study [16]. The results indicate that this set of social and human factors are perceived as influencing SDT productivity. However, the magnitude of this influence is still unknown because non-technical factors are difficult to measure [6].

Finally, considering the importance of Human Factors in Software Development, this topic can be considered a subfield of Empirical Software Engineering [17], where psychological knowledge plays an important role [6].

3 Systematic Mapping Protocol

Systematic mapping is a particular type of literature review focused on understanding the behavior of a field of knowledge at the research level and the main challenges that are still to be solved [18]. This mapping considers both primary and secondary research, and adopts the protocol used by Brereton et al. [19], which includes three phases:

- Phase 1: study planning, where research questions, scope, research criteria, and study selection criteria are defined.
- Phase 2: execution of the study, which includes the selection and classification of studies.
- Phase 3: analysis of results and response to each of the research questions.

3.1 Phase 1: Study Planning

Planning the study involves defining the research questions, scope, research criteria, and selection criteria. According to the object of study, the systematic mapping is focused on characterizing the scientific production on measurement of human aspects in software development. Therefore, the research questions are:

- RQ1: What human aspects are measured in Software Engineering?
- RQ2: What are the sources of data used to quantify human aspects in software engineering?
- RQ3: What data analysis methods are used in Software Engineering to analyze data related to human aspects?

The limits of the research according to the guidelines proposed by Petticrew and Roberts [20], the population of interest are people who are part of software development teams or companies without considering a specific geographic location. Research published as of January 2010, inclusive, was considered, and the review of the selected studies was conducted by an expert in SE and another in engineering research.

The search for publications was performed in the IEEE Xplore, ACM, Science Direct, and Springer databases to cover the objective of the systematic mapping. The terms that can account for the topic of interest are related to "measurement instruments" or "evaluation" of "human factors" or "human aspects" in Software Engineering or Development, so the search string used was:

(assessment OR "measuring instruments") AND ("human factor" OR "human aspect" OR "soft skill") AND ("software engineering" OR "software development").

The term "human aspect" was used because it is an area of research in software engineering, related to human resource management [21]. This area includes the analysis of "soft skill" that is usually related to "human factor" [6]. On the other hand, the term "assessment" was used because it is widely used in clinical psychology [22], while the term "measuring instrument" is broader [23].

The study classification procedure began with a review of the title, keywords, and abstract of each candidate study to identify the relationship with the topic of interest. In some cases, it was necessary to review other sections such as the introduction and conclusions, when the previous ones were not conclusive to identify the respective relationship with the topic of interest.

It is necessary to define the criteria required to make decisions regarding the inclusion or exclusion of studies in the systematic mapping to select the candidates that will be part of the initial sample of studies. Table 1 presents the inclusion and exclusion criteria considered in this mapping and used to select the studies.

Table 1. Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
- Studies involving human aspects in software engineering - Studies published since 2010	- Studies without complete available document - Studies not written in English - Duplicate investigations

3.2 Phase 2: Execution of Study

The execution of the study is the mapping phase that allows the aspects planned in the previous phase to be implemented. That is, to search according to the search string defined in the selected databases and based on the inclusion and exclusion criteria, to select the studies and then classify them.

The execution of the search string in the identified databases yielded an initial capture of 1533 studies. The application of the established selection criteria and the review of the general elements of each document (title, abstract and keywords, and introduction and conclusion, when necessary), led to the selection of 99 studies, representing 6.46% of the total number of studies reported in the initial capture (Fig. 1).

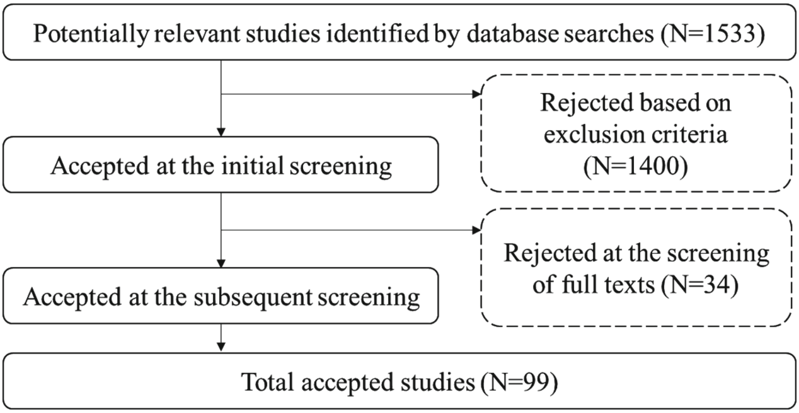

Potentially relevant studies identified by database searches (N=1533)
Rejected based on exclusion criteria (N=1400)
Accepted at the initial screening
Rejected at the screening of full texts (N=34)
Accepted at the subsequent screening
Total accepted studies (N=99)

Fig. 1. Flow diagram of the data collection procedure

The number of studies captured and selected in each database is presented in Table 2. Relevant information on the 99 selected publications can be found at https://github.com/lmrestrepo/HA-SD.git.

Table 2. Number of studies by database consulted

	Science Direct	Springer	IEEE	ACM	Total
Captured studies	623	473	250	187	1533
Selected studies	35	18	10	36	99

Once the studies of interest were selected, they were classified according to a) year of publication, b) country of affiliation of the authors, c) type of study (case study, observation, experiment, document analysis), d) human aspects treated as dependent variables, e) human aspects treated as independent variables, f) measurement instruments used, and g) data analyzing methods.

4 Phase 3: Analysis of Results

This section presents the results obtained by classifying the selected studies. This phase aims to answer the three questions posed in the planning phase of this systematic mapping.

4.1 General Results of the Study

Figure 2 shows the behavior of studies by year considering publications in journals and academic events. There is an increasing trend of studies related to the measurement of human aspects in SE between 2011 and 2014. However, from 2014 onwards, an oscillatory behavior is presented suggesting a stabilization of the topic of interest for the last 7 years.

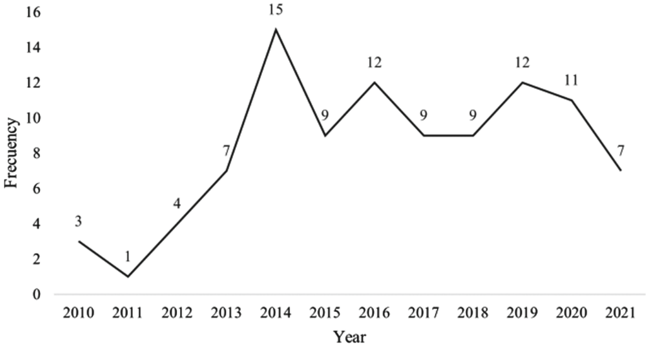


Fig. 2. Number of studies per year

Table 3 lists the first five of each type of publication, according to the number of studies that were selected. In total, 54 studies in journals (54.55%) and 45 studies in academic events (45.45%) were considered.

Table 3. Number of studies by type of publication

Journal name	#	Proceeding name	#
Information and Software Technology	19	CHASE Cooperative and Human Aspects of Software Engineering	7
The Journal of Systems and Software	9	EASE Evaluation and Assessment in Software Engineering	6
Empirical Software Engineering	8	ESEM Empirical Software Engineering and Measurement	5
Computers in Human Behavior	3	ICSE International Conference on Software Engineering	4
Transactions on Computing Education	2	EuroSPI Systems, Software and Services Process Improvement	2
Transactions on software engineering	2	SIGSOFT Software Engineering Notes	2

According to the countries of affiliation of the authors who have participated in the studies selected for this systematic mapping, 19.61% of the total number of authors reported are from Brazil, 8.12% are from Sweden, 6.44% are from Spain and 5.60% are from the United States. The continent with the most researchers working on the measurement of human aspects in SE is Europe with 46.78%, followed by South America (20.45%), and Asia (14.01%).

The survey study stands out (55 studies - 55.56%), where information is obtained from individuals through a questionnaire or an interview and then processed quantitatively or qualitatively, depending on the type of questions and the interest of the research. Some studies are related to literature reviews (17 studies - 17.17%).

Case studies (13 studies - 13.13%) are less frequent, but they are applied when it is desired to describe the behavior of an individual or a reduced set of individuals. Publications were found that rely on the analysis of documents (8 studies - 8.08%) such as organizational repositories and e-mails to study human aspects.

Studies were also found in which experiments or quasi-experiments were carried out in which some variables were controlled (3 studies - 3.03%). Finally, three laboratory studies were found in which electronic devices or sensors are used to collect data and, therefore, are performed under specific working conditions.

4.2 What Human Aspects are Measured in Software Engineering?

We founded research in which cause-effect relationships of human aspects with other variables were studied, where human aspects can be treated as dependent (explained) variables or as independent (explanatory) variables. In cases where no cause-effect relationships were studied, the variables were taken as independent variables.

Table 4 presents the variables and the number of studies associated with these variables (number greater than or equal to 2), classified as independent and dependent. In some studies, an aspect can be considered as an independent variable and in others as a dependent variable, as occurs with decision-making and motivation.

Table 4. Independent and dependent variables

Independent variable	#	Dependent variable	#
Personality	29	Performance	11
Experience	5	Team building	5
Communication	4	Productivity	5
Social interaction	3	Motivation	4
Team size	2	Quality	4
Commitment	2	Job satisfaction	3
Emotional intelligence	2	Yield	3
Trust	2	Team climate	2
Task characteristics	2	Decision making	2
Decision making	2	Role conflict	2
Motivation	2	Project success	2
Leadership	2		
Emotion	2		
Team autonomy	2		
Happiness	2		

As evidenced in Table 4, the human aspect that has been most studied in SE is personality, which is related to the fact that the software development process is predominantly social [24] and people-centered activity [17].

The literature reviews have been carried out regarding personality. One of the studies is related to the identification of personality aspects that have been of interest in SE [25]. Others have studied the relationship between personality and performance [26, 27] and team climate [28, 29]. One literature review studied the relationship between personality and decision making [30] and another compared the instruments used in SE to measure personality [31].

Personality has been included as an independent variable in research that studies cause-effect relationships. Table 5 consolidates the variables that have been related and the respective references. There is interest in studying how personality influences efficiency and how to build work teams and assign tasks to optimize the software development process.

Table 5. Studies related to personality like independent variable

Dependent variable	References
Performance, productivity, yield	[32–36]
Decision making	[37]
Communication	[38]
Team climate	[39, 40]
Team building, role assignment, task preference	[21, 41–49]
Attitude and behavior	[50–52]
Trust to reuse code	[53]

Experience is another human aspect that has been taken into account to estimate performance [32, 54]. Juneja [55] proposes an instrument to measure the performance of programmers based on experience among other aspects. Also, the relationship between experience and autonomy has been studied [56] and between experience and team process background [57].

Considering the 13 social and human factors that are perceived to influence SDT productivity [15], some of them are directly evidenced in this mapping and are related in Table 6.

One of the aspects that have been studied as a dependent variable is motivation because it favors efficiency in software development projects [74]. Also, factors that influence motivation have been studied [75], and one of the studies analyzed has focused particularly on the motivation of the engineers in charge of the tests [72]. In addition, a study was identified that proposes a way to measure motivation through sensors [76], and another study where the relationship between motivation and job satisfaction was studied [66, 69].

Job satisfaction is another human aspect of interest for researchers, and according to the above, its relationship with motivation. This mapping identified an instrument built

Table 6. Studies related to human aspects that are perceived as influencing productivity

Independent variable	Dependent variable	References
Communication	Project performance and success	[58–61]
Social interaction (interpersonal relationships)	Software quality, team tacit knowledge acquisition, and value creation	[62–64]
Emotional intelligence	Task preference and team performance	[43, 65]
Motivation	Job satisfaction and developer skills	[66, 67]
Commitment	Team building, motivation, and job satisfaction	[68, 69]
Leadership	Project success and team process history	[57, 59]
Collaboration	Team building and teamwork quality	[68, 70]
Autonomy (team)	Project results and quality of teamwork	[70, 71]
Cohesion	Quality of teamwork	[70]
Innovation (creativity)	Motivation	[72]
Job satisfaction	Turnover	[73]

to identify whether the clarity of equipment standards and psychological safety impact job satisfaction and SDT performance [77].

Team climate is an aspect that has attracted the attention of researchers. Soomro et al. [28] conducted a systematic literature review on personality traits that influence the climate of the SDT and, subsequently, Vishnubhotla et al. [40] build a regression model between both variables in agile teams.

Finally, decision-making has also been studied. Freitas et al. [30] did a literature review on the personality of decision-makers in SE, and then presented a regression model between both aspects [37].

4.3 What are the Sources of Data Used to Quantify Human Aspects in Software Engineering?

Table 7 presents the data sources most frequently used in the studies analyzed. Out of 176 records, considering that several sources can be used for an investigation, 22.73% correspond to questionnaires and 9.09% to interviews. Researchers can use instruments designed by themselves, use instruments designed and validated by other authors (e.g., psychometric tests), or make adaptations of instruments used in previous research. In this mapping, any of the three options is considered as ‘Questionnaire’.

Table 7. Number of studies by type of data source

Data source	#	Data source	#
Questionnaire	40	Organizational data	2
Interview	16	Self-Assessment Manikin	2
International Personality Item Pool	11	Maslach Burnout Inventory	2
Big Five scores	7	Scale of Positive and Negative Experience	2
Observation	4	Document analysis	2
Myers–Briggs Type Indicator	3	Team Climate Inventory	2
Electronic device	3	GitHub repository	2
Jazz repository (IBM)	3	Work Design Questionnaire	2

Surveys, case studies, and experiments are the main methods of empirical research in SE [78]. The surveys, supported by questionnaires and interviews, are the data sources that are most used in the investigations of this mapping.

Both the questionnaire and the interviews are empirical research tools used to collect information about the processes and skills of software developers through self-report [7]. In several investigations, they are used simultaneously, to obtain additional information that can be obtained using only one of these tools [42, 56, 79–84].

Concerning surveys (questionnaires and interviews), some limitations should be considered. One of these is to ensure that the sample size is representative of the population to be studied. The sample size is necessary to achieve the generalization of the findings. This limitation is due to its requirement in the identification of the unit of analysis [85]. In addition, Kitchenham et al. [86] recommend reporting the response rate and, presenting how study participants were recruited and selected, which may constitute an additional constraint for the management and administration of information and the collection process.

Instruments that have been designed and validated by other authors, employed iteratively in research, and which are usually called acronyms, are considered formalized (some even with restricted use by licensing), as in the case of personality instruments, team climate, job burnout, among others. Table 8 presents the formalized instruments found in this mapping, the aspect they measure, and the references of the studies in which they were used.

Table 8. Formalized instruments

Formalized instrument	Aspect	References
International Personality Item Pool (IPIP)	Personality	[33–37, 40, 43, 44, 46, 50, 53]
Big Five scores	Personality	[32, 35, 38, 39, 47, 51, 52]
Myers–Briggs Type Indicator – MBTI	Personality	[21, 45, 68]
Self-Assessment Manikin (SAM)	Visual stimuli	[83, 87]
Maslach Burnout Inventory (MBI)	Job Burnout	[88, 89]
Scale of Positive and Negative Experience	Feelings	[90, 91]
Team Climate Inventory (TCI)	Team climate	[39, 40]
Work Design Questionnaire (WDQ)	Job characteristics	[88, 89]
Multidimensional Work Motivation Scale	Laboral motivation	[92]
Belbin Self-Perception Test	Team roles	[68]
Team Tacit Knowledge Measure (TTKM)	Tacit knowledge of the team	[63]
Multifactor Leadership Questionnaire (MLQ)	Leadership	[59]
Positive And Negative Affect Schedule	Affect	[60]
Intrinsic Motivation Inventory (IMI)	Intrinsic motivation	[67]
Team Selection Inventory (TSI)	Team selection	[39]
Keirsey Temperament Sorter (KTS)	Temperament	[34]

Psychometric instruments are easy to apply and score and allow measuring the characteristics of interest of a person at a given time. However, the results can be affected by the individual's temporal events, the inherent interaction between the examiner and the examinee, and the possibility of distorting or simulating the examinee's responses [93].

In line with the above, Andersson et al. [94] made a comparison between the ratings given by behavioral observers and the self-evaluation ratings to measure team performance. In this comparison, from a quasi-experiment, they found that observation-based techniques are more reliable than those based on self-assessment. The observation allows corroborating the relationship between what the teams say they do and what they do [95, 96]. In addition, it allows for identifying the strategies that individuals use when facing stressful situations [80].

As an alternative to psychometric testing, the use of electronic devices has been proposed. Bordel and Alcarria [76] designed a device to automatically evaluate motivation in Industry 4.0 scenarios with Environmental Intelligence infrastructure. This device is based on body and environmental sensors that account for the physiological and emotional signals of the individual. Another proposal is that reported by Girardi et al. [87], who used an electronic device to identify the emotions of software developers while working on it. Also, Fritz et al. [97] report the use of psychophysiological sensors to detect when software developers are facing a difficult task and prevent them from making mistakes.

4.4 What Data Analysis Methods are Used in Software Engineering to Analyze Data Related to Human Aspects?

Table 9 presents the main data analysis methods reported in the studies selected in this mapping to analyze the data obtained related to human aspects, excluding those of a philosophical and opinion type. In total, 25 data analysis methods were reported, but this table only presents those that had a frequency of use greater than or equal to two.

Table 9. Data analysis methods

Method	#	Method	#
Descriptive statistic	48	Normality test	6
Correlation analysis	27	Cluster analysis	4
Regression analysis	22	Data mining	3
Mean comparison study	20	Linguistic analysis	3
Qualitative analysis	17	Machine learning	2
Factor analysis	11	Archetypal analysis	2
Reliability calculation (Cronbach’s Alpha)	11	Bayesian statistic	2
Principal component analysis	7		

A recent systematic review of the literature indicates that the statistical methods most commonly used in SE are descriptive statistics, power analysis of statistical tests, the goodness of fit tests, parametric and non-parametric tests, error type I, confidence intervals, analysis of latent variables and finally, practical significance and size of the effect [98]. All the methods mentioned are confirmed in the set of studies on this mapping.

The predominant method is descriptive statistics, used in 60.00% of the studies where data analysis is possibly required. This method is usually used in the first steps of statistical analysis because it allows summarizes relevant information [98, 99] and usually includes measures of central tendency and dispersion [100]. This method is used in some studies to present preliminary results [45, 49, 55, 66, 101, 102], and in other studies is before the use of more advanced statistical methods [33, 34, 40, 42, 43, 47, 50, 52, 54, 57, 60, 63, 67, 70, 87–90, 103–106].

A recurrent analysis method is correlation analysis, used in 33.75% of the studies analyzed in this mapping in which data analysis was possibly required. These studies in some cases can lead to regression analysis (27.50%), and subsequently to the comparison of means (25.00%). Normality testing is required to do a Pairwise Comparison, but only six of the studies reported this. Regression models allow modeling a variable according to others and are usually accompanied by correlation studies [107]. These models have been used to estimate programming performance based on personality and cognitive styles [33]. Information is also reported where regression models are used to estimate decision-making from personality [37] and to estimate performance based on social interaction, transactive memory, and tacit team knowledge [63], among others.

Concerning studies that present designed instruments, they usually support their research in factor analysis (13.75%). On some occasions, it was accompanied by principal component analysis as an estimation method (8.75%). Feldt et al. [105] designed an instrument to measure willingness and openness to organizational change based on participation, knowledge, and the need for change. Marsicano et al. [57] propose a questionnaire called Teamwork Process Antecedents (TPA). This questionnaire aims to measure the background of equipment processes for use in research and the management of equipment in practice. In both cases, Cronbach's Alpha is reported as a measure of reliability. For its part, Gren [108] makes some recommendations to ensure the construct validity and reliability of the instruments, taking into account the stability and internal consistency. Likewise, Lloret-Segura et al. [109] present a guide for doing exploratory factor analysis.

Given the recent interest in analyzing human aspects in SE, which have a latent nature, it is necessary to use methods that work with this type of variables, as is the case of models of structural equations of partial least squares [110]. Xiang et al. [65] used structural equations to study the relationship between emotional intelligence and the shared mental model, and between the shared mental model and team performance in the analysis of requirements. Basirati et al. [81] used this method for to understand the impact of conflict on the success of software development projects for different types of conflicts and different environments.

Naturally, the use of analytical methods is not unrelated to software development [111]. The use of machine learning allows to calculate the probability of replacement of a person and resignation in a software development company [73] and, can be the basis for defining classifiers to predict difficult tasks [97]. For its part, data mining is a useful method when it is necessary to analyze large volumes of data [38, 68, 73].

The analysis method depends on the type of data being analyzed. Several studies present their results based on a qualitative analysis because of the use of open-ended interviews. In particular, semi-structured interviews allow flexibility but require intelligence, sensitivity, preparation, and agility on the part of the interviewer [112]. This method is used when the interest is to know the individual opinion of people who are part of a group [42, 56, 58, 69, 80, 81, 83, 84, 113–115]. In some cases, researchers prefer to use structured interviews with previously defined questions [70, 79, 95, 116].

5 Implications for the Practice and Research

The results of this systematic mapping suggest that human aspects such as personality and motivation have an impact on the performance of the SDT, so its measurement is a relevant aspect in organizations. Personality is an aspect to be considered for recruiting, selecting, and retaining talent, as well as for forming work teams. Concerning motivation, organizations should identify the aspects that promote it and, on that basis, verify the need to propose strategies to reach the desired levels.

Now, the list of human aspects is broad, and several of them have also been identified as influential in the performance of the teams. Therefore, organizations have at their disposal the tools designed and tested to be included in their management processes.

Concerning training programs in SE, training in soft skills is fundamental because it has been shown to influence the success of projects. For technical training, the fact that knowledge in statistics and data analysis methods is required to support studies of interdependence where human aspects are involved is highlighted.

The research on human aspects has been oriented to the study of the personality of the members of the SDT, in their relationship with performance and efficiency. It has also been analyzed how to take advantage of personality to form teams, although it is an area still to explore. Likewise, interest in other human aspects such as motivation, communication, or collaboration in SE has been identified. Recognizing the human aspects that are of interest in the area and that have not been studied exhaustively, represents an opportunity for future research.

According to the results, human aspects have traditionally been measured using psychometric tests, which can give biased results when situational. Therefore, in this aspect of the results, there is a challenge to be solved related to the use of measurement mechanisms. Mechanisms are needed to ensure consistency between what software developers say they do and what they do.

Finally, the use of statistical methods is a common practice for studies related to instrument design and for addressing analysis. These methods aim to identify relationships of interdependence, maintaining in all cases the mathematical rigor and data processing. However, emerging data analysis methods are identified that can respond to the needs of data analysis in this context and become complementary methods to enrich knowledge around work of this research.

5.1 Study Limitations

This study focused on scientific databases such as 1) Science Direct, 2) Springer, 3) IEEE and 4) ACM. A limitation associated with these databases is related to the possible exclusion of relevant studies on the measurement of human aspects in SE, published in other databases with less coverage. However, the four databases mentioned were selected for their wide scope in the areas of engineering, SE, and Computer Science, as well as their frequency of use among the scientific community as a source of recent and reliable information in the area.

This systematic mapping was carried out with 99 studies, published in journals or academic events. This amount may vary by including terms more specific to those used in this study (human aspects, human factors, soft skill). However, the sample of studies selected in this mapping made it possible to identify the aspects towards which the research has been oriented and to identify future work opportunities in the area.

6 Conclusions

This paper presents a systematic mapping that includes the review of 99 scientific studies related to the human aspects that have been of interest to measure in SE between 2010 and 2021. We analyzed the number of publications per year, by type of document, by country of affiliation of authors, by type of research, and by type of study.

Personality is the most studied human aspect in SE as an independent variable, mainly to explain the performance of SDT and to assign roles and form teams. However, there is a broad spectrum of human aspects that has already attracted the interest of some researchers, and where there are opportunities to delve deeper.

Human aspects are variables that are measured indirectly from other observable variables, so their latent nature allows surveys, through questionnaires and interviews, to be the most frequent source of data. The use of instruments based on self-reporting is frequent and, therefore, the results depend on the conditions under which the data were captured. Therefore, adopting methodologies that allow taking data that ensure consistency between what is done and what is said to be done is a challenge to be addressed.

Descriptive statistics was the most common method of data analysis, either to present preliminary results or as a method prior to others such as regression analysis or factor analysis. Statistical methods predominate in the analysis of the data obtained in these investigations, without forgetting the contribution of qualitative analysis when interviews are conducted. In any case, the usual practice of using several methods of analysis in the same study will depend on the nature of the data obtained.

This mapping suggests that the study of human aspects in HE is a topic of interest and is still under construction because there are several aspects that can be further explored. A systematic literature review for each aspect can clarify specific lines of work. Additionally, estimating the influence of human aspects on the efficiency, performance, or productivity of the SDT would allow designing intervention strategies focused on these aspects so that it is possible to improve the management of human talent in projects.

Future work is related to the use of specialized techniques in modeling the context as a dynamic system to determine the relationship between the different factors. One line of future work is related to broadening the scope of this research to include dimensions of social and human factors such as those related to happiness, stress management or anxiety, among others.

References

1. Staron, M., Meding, W.: Software development measurement programs: development, management and evolution. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91836-5
2. Hazzan, O., Hadar, I.: Why and how can human-related measures support software development processes? J. Syst. Softw. **81**, 1248–1252 (2008). https://doi.org/10.1016/j.jss.2008.01.037
3. Acuña, S.T., Gómez, M., Juristo, N.: How do personality, team processes and task characteristics relate to job satisfaction and software quality? Inf. Softw. Technol. **51**, 627–639 (2009). https://doi.org/10.1016/j.infsof.2008.08.006
4. Sadowski, C., Zimmermann, T. (eds.): Rethinking Productivity in Software Engineering. Apress, Berkeley, CA (2019). https://doi.org/10.1007/978-1-4842-4221-6
5. Machuca-Villegas, L., Gasca-Hurtado, G.P.: Towards a social and human factor classification related to productivity in software development teams. In: Mejia, J., Muñoz, M., Rocha, Á., A. Calvo-Manzano, J. (eds.) CIMPS 2019. AISC, vol. 1071, pp. 36–50. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-33547-2_4
6. Capretz, L.F., Ahmed, F., da Silva, F.Q.B.: Soft sides of software. Inf. Softw. Technol. **92**, 92–94 (2017). https://doi.org/10.1016/j.infsof.2017.07.011
7. Cha, S., Taylor, R.N., Kang, K.: Handbook of software engineering. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-00262-6
8. Schmidt, R.F.: Software engineering fundamentals. In: Software Engineering, pp. 1–6 (2013)
9. Licorish, S.A., Macdonell, S.G.: Exploring software developers' work practices: task differences, participation, engagement, and speed of task resolution. Inf. Manag. **54**, 364–382 (2017). https://doi.org/10.1016/j.im.2016.09.005
10. Sudhakar, G.P., Farooq, A., Patnaik, S.: Soft factors affecting the performance of software development teams. Team Perform. Manag. **17**, 187–205 (2011). https://doi.org/10.1108/13527591111143718
11. Trendowicz, A., Münch, J.: Factors influencing software development productivity-state-of-the-art and industrial experiences. In: Advances in Computers, pp. 185–241 (2009)
12. Canedo, E.D., Santos, G.A.: Factors affecting software development productivity: an empirical study. In: ACM International Conference Proceeding Series, pp. 307–316. Association for Computing Machinery (2019)
13. Devadas, U.M., Dharmapala, Y.Y.: Soft skills evaluation in the information technology and business process management industry in Sri Lanka: skills, methods and problems. Int. J. Econ. Bus. Hum. Behav. **2**(3) (2021). https://doi.org/10.5281/zenodo.5280309
14. Ahmed, F., Capretz, L.F., Campbell, P.: Evaluating the demand for soft skills in software development. IT Prof. **14**, 44–49 (2012). https://doi.org/10.1109/MITP.2012.7
15. Machuca-Villegas, L., Gasca-Hurtado, G.P., Restrepo-Tamayo, L.M., Morillo Puente, S.: Social and human factor classification of influence in productivity in software development teams. In: EuroSPI Systems, Software and Services Process Improvement, pp. 717–729 (2020)
16. Machuca-Villegas, L., Gasca-Hurtado, G.P., Morillo Puente, S., Restrepo-Tamayo, L.M.: An instrument for measuring perception about social and human factors that influence software development productivity. J. Univers. Comput. Sci. **27**, 111–134 (2021). https://doi.org/10.3897/jucs.65102
17. Amrit, C., Daneva, M., Damian, D.: Human factors in software development: on its underlying theories and the value of learning from related disciplines. A guest editorial introduction to the special issue. Inf. Softw. Technol. **56**, 1537–1542 (2014). https://doi.org/10.1016/j.infsof.2014.07.006

18. Petersen, K., Feldt, R., Mujtaba, S., Mattsson, M.: Systematic mapping studies in software engineering. In: 12th International Conference on Evaluation and Assessment in Software Engineering, EASE 2008 (2008)
19. Brereton, P., Kitchenham, B.A., Budgen, D., Turner, M., Khalil, M.: Lessons from applying the systematic literature review process within the software engineering domain. J. Syst. Softw. **80**, 571–583 (2007). https://doi.org/10.1016/j.jss.2006.07.009
20. Petticrew, M., Roberts, H.: Systematic Reviews in the Social Sciences: A Practical Guide (2008)
21. Varona, D., Capretz, L.F.: A comparison of junior and senior software engineering students' personalities. In: CHASE Cooperative and Human Aspects of Software Engineering, pp. 131–132 (2014)
22. Meyer, G.J., et al.: Psychological testing and psychological assessment. Am. Psychol. **56**, 128–165 (2001). https://doi.org/10.1037//OOO3-O66X.56.2.128
23. Gawronski, B., De Houwer, J.: Implicit measures for social and personality psychology. In: Handbook of Research Methods in Social and Personality Psychology, pp. 282–310 (2014)
24. Tenenberg, J.: An institutional analysis of software teams. Int. J. Hum. Comput. Stud. **66**, 484–494 (2008). https://doi.org/10.1016/j.ijhcs.2007.08.002
25. Cruz, S., Fabio, Q.B., Fernando, L.: Forty years of research on personality in software engineering: a mapping study. Comput. Human Behav. **46**, 94–113 (2015). https://doi.org/10.1016/j.chb.2014.12.008
26. Ferreira, N.N.V., Langerman, J.J.: The correlation between personality type and individual performance on an ICT project. In: ICCSE International Conference on Computer Science & Education, pp. 425–430 (2014)
27. Wiesche, M., Krcmar, H.: The relationship of personality models and development tasks in software engineering. In: SIGMIS-CPR Conference on Computers and People Research, pp. 149–161 (2014)
28. Soomro, A.B., Salleh, N., Mendes, E., Grundy, J., Burch, G., Nordin, A.: The effect of software engineers' personality traits on team climate and performance: a systematic literature review. Inf. Softw. Technol. **73**, 52–65 (2016). https://doi.org/10.1016/j.infsof.2016.01.006
29. Bano, A., Salleh, N., Mendes, E., Grundy, J., Burch, G., Nordin, A.: The effect of software engineers' personality traits on team climate and performance: a systematic literature review. Inf. Softw. Technol. **73**, 52–65 (2016). https://doi.org/10.1016/j.infsof.2016.01.006
30. Freitas, F., Mendes, E., Salleh, N.: The relationship between personality and decision-making: a systematic literature review. Inf. Softw. Technol. **111**, 50–71 (2019). https://doi.org/10.1016/j.infsof.2019.03.010
31. Jia, J., Zhang, P., Zhang, R.: A comparative study of three personality assessment models in software engineering field. In: ICSESS International Conference on Software Engineering and Service Science, pp. 7–10. IEEE (2015)
32. Hannay, J.E., Arisholm, E., Engvik, H., Sjøberg, D.I.K.: Effects of personality on pair programming. IEEE Trans. Softw. Eng. **36**, 61–80 (2010)
33. Huang, F., Liu, B., Song, Y., Keyal, S.: Science of computer programming the links between human error diversity and software diversity: implications for fault diversity seeking. Sci. Comput. Program. **89**, 350–373 (2014). https://doi.org/10.1016/j.scico.2014.03.004
34. Gulati, J.: A study of relationship between performance, temperament and personality of a software programmer. SIGSOFT Softw. Eng. Notes **41**, 1–5 (2016). https://doi.org/10.1145/2853073.2853089
35. Karimi, Z., Baraani-Dastjerdi, A., Ghasem-Aghaee, N., Wagner, S.: Links between the personalities, styles and performance in computer programming. J. Syst. Softw. **111**, 228–241 (2016). https://doi.org/10.1016/j.jss.2015.09.011

36. Caulo, M., Francese, R., Scanniello, G., Tortora, G.: Relationships between personality traits and productivity in a multi-platform development context. In: The International Conference on Evaluation and Assessment in Software Engineering (EASE), pp. 70–79 (2021)
37. Mendes, F., Mendes, E., Salleh, N., Oivo, M.: Insights on the relationship between decision-making style and personality in software engineering. Inf. Softw. Technol. **136**, 106586 (2021). https://doi.org/10.1016/j.infsof.2021.106586
38. Licorish, S.A., Macdonell, S.G.: Communication and personality profiles of global software developers. Inf. Softw. Technol. **64**, 113–131 (2015). https://doi.org/10.1016/j.infsof.2015.02.004
39. Gómez, M.N., Acuña, S.T.: A replicated quasi-experimental study on the influence of personality and team climate in software development. Empir. Softw. Eng. **19**(2), 343–377 (2013). https://doi.org/10.1007/s10664-013-9265-9
40. Vishnubhotla, S.D., Mendes, E., Lundberg, L.: Investigating the relationship between personalities and agile team climate of software professionals in a telecom company. Inf. Softw. Technol. **126**, 106335 (2020). https://doi.org/10.1016/j.infsof.2020.106335
41. Martinez, L., Guillermo, L., Rodríguez-Díaz, A., Castro, J.: Experiences in software engineering courses using psychometrics with RAMSET. In: ITiCSE Innovation and Technology in Computer Science Education, pp. 244–248 (2010)
42. Silva, F.Q.B., et al.: Team building criteria in software projects: a mix-method replicated study. Inf. Softw. Technol. **55**, 1316–1340 (2013). https://doi.org/10.1016/j.infsof.2012.11.006
43. Kosti, M., Feldt, R., Angelis, L.: Personality, emotional intelligence and work preferences in software engineering: an empirical study. Inf. Softw. Technol. **56**, 973–990 (2014). https://doi.org/10.1016/j.infsof.2014.03.004
44. Papatheocharous, E., Belk, M., Nyfjord, J., Germanakos, P., Samaras, G.: Personalised continuous software engineering. In: RCoSE Rapid Continuous Software Engineering, pp. 57–62 (2014)
45. Capretz, F., Varona, D., Raza, A.: Influence of personality types in software tasks choices. Comput. Human Behav. **52**, 373–378 (2015). https://doi.org/10.1016/j.chb.2015.05.050
46. Kosti, M.V., Feldt, R., Angelis, L.: Archetypal personalities of software engineers and their work preferences: a new perspective for empirical studies. Empir. Softw. Eng. **21**(4), 1509–1532 (2015). https://doi.org/10.1007/s10664-015-9395-3
47. Yilmaz, M., Connor, R.V.O., Colomo-Palacios, R., Clarke, P.: An examination of personality traits and how they impact on software development teams. Inf. Softw. Technol. **86**, 101–122 (2017). https://doi.org/10.1016/j.infsof.2017.01.005
48. Muñoz, M., Peña, A., Mejia, J., Rangel, N., Torres, C., Hernández, L.: Building high effectives teams using a virtual environment. In: EuroSPI Systems, Software and Services Process Improvement, pp. 554–568 (2018)
49. Akarsu, Z., Orgun, P., Dinc, H., Gunyel, B., Yilmaz, M.: Assessing personality traits in a large scale software development company: exploratory industrial case study. In: Walker, A., O'Connor, R.V., Messnarz, R. (eds.) EuroSPI 2019. CCIS, vol. 1060, pp. 192–206. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-28005-5_15
50. Feldt, R., Angelis, L., Torkar, R., Samuelsson, M.: Links between the personalities, views and attitudes of software engineers. Inf. Softw. Technol. **52**, 611–624 (2010). https://doi.org/10.1016/j.infsof.2010.01.001
51. Calefato, F., Lanubile, F., Vasilescu, B.: A large-scale, in-depth analysis of developers' personalities in the Apache ecosystem. Inf. Softw. Technol. **114**, 1–20 (2019). https://doi.org/10.1016/j.infsof.2019.05.012
52. Licorish, S.A., Macdonell, S.G.: Personality profiles of global software developers. In: EASE Evaluation and Assessment in Software Engineering, pp. 1–10 (2014)

53. Capiola, A., et al.: Trust in software: attributes of computer code and the human factors that influence utilization metrics. In: Stephanidis, C. (ed.) HCII 2019. CCIS, vol. 1032, pp. 190–196. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-23522-2_24
54. Dieste, O., et al.: Empirical evaluation of the effects of experience on code quality and programmer productivity: an exploratory study. Empir. Softw. Eng. **22**(5), 2457–2542 (2017). https://doi.org/10.1007/s10664-016-9471-3
55. Juneja, K.: Design of programmer's skill evaluation metrics for effective team selection. Wireless Pers. Commun. **114**(4), 3049–3080 (2020). https://doi.org/10.1007/s11277-020-07517-6
56. Da Silva, L.M., et al: Autonomy in software engineering: a preliminary study on the influence of education level and professional experience. In: 2017 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement Autonomy, pp. 229–234 (2017)
57. Marsicano, G., Silva, F.Q.B., Seaman, C.B., Adaid-Castro, B.G.: The Teamwork Process Antecedents (TPA) questionnaire: developing and validating a comprehensive measure for assessing antecedents of teamwork process quality. Empir. Softw. Eng. **25**, 3928–3976 (2020)
58. Fagerholm, F., Ikonen, M., Kettunen, P., Münch, J., Roto, V., Abrahamsson, P.: Performance alignment work: how software developers experience the continuous adaptation of team performance in Lean and Agile environments. Inf. Softw. Technol. **64**, 132–147 (2015). https://doi.org/10.1016/j.infsof.2015.01.010
59. Van Kelle, E., Visser, J., Plaat, A., van der Wijst, P.: An empirical study into social success factors for agile software development. In: CHASE Cooperative and Human Aspects of Software Engineering, pp. 77–80 (2015)
60. Schneider, K., Liskin, O., Paulsen, H., Kauffeld, S.: Media, mood, and meetings: related to project success? ACM Trans. Comput. Educ. **15**, 1–33 (2015)
61. de Sá Leitão Júnior, N.G., de Farias Junior, I.H., Marczak, S., Santos, R., Furtado, F., de Moura, H.P.: Evaluation of a preliminary assessment method for identifying the maturity of communication in distributed software development (2017)
62. Bettenburg, N., Hassan, A.E.: Studying the impact of social interactions on software quality. Empir. Softw. Eng. **18**, 375–431 (2013). https://doi.org/10.1007/s10664-012-9205-0
63. Ryan, S., Connor, R.V.O.: Acquiring and sharing tacit knowledge in software development teams: an empirical study. Inf. Softw. Technol. **55**, 1614–1624 (2013). https://doi.org/10.1016/j.infsof.2013.02.013
64. Alahyari, H.: The role of social interactions in value creation in agile software development processes. In: SSE Social Software Engineering, pp. 17–20 (2015)
65. Xiang, C., Yang, Z., Zhang, L.: Improving IS development teams' performance during requirement analysis in project—the perspectives from shared mental model and emotional intelligence. Int. J. Proj. Manag. **34**, 1266–1279 (2016). https://doi.org/10.1016/j.ijproman.2016.06.009
66. França, C., Sharp, H., Silva, F.Q.B.: Motivated software engineers are engaged and focused, while satisfied ones are happy. In: ESEM Empirical Software Engineering and Measurement, pp. 1–8 (2014)
67. Kuusinen, K., Petrie, H., Fagerholm, F., Mikkonen, T.: Flow, intrinsic motivation, and developer experience in software engineering. In: International Conference on Agile Software Development, pp. 104–117 (2016)
68. André, M., Baldoquín, M.G., Acuña, S.T.: Formal model for assigning human resources to teams in software projects. Inf. Softw. Technol. **53**, 259–275 (2011). https://doi.org/10.1016/j.infsof.2010.11.011
69. França, C., Silva, F.Q.B., Sharp, H.: Motivation and satisfaction of software engineers. IEEE Trans. Softw. Eng. **46**, 118–140 (2020)

70. Freire, A., Perkusich, M., Saraiva, R., Almeida, H., Perkusich, A.: A Bayesian networks-based approach to assess and improve the teamwork quality of agile teams. Inf. Softw. Technol. **100**, 119–132 (2018). https://doi.org/10.1016/j.infsof.2018.04.004
71. Günsel, A., Açikgšz, A., Tükel, A., Öğüt, E.: The role of flexibility on software development performance: an empirical study on software development teams. Procedia - Soc. Behav. Sci. **58**, 853–860 (2012). https://doi.org/10.1016/j.sbspro.2012.09.1063
72. Santos, R.E.S., Correia-Neto, J.S., Silva, F.Q.B., Souza, R.E.C.: Would you like to motivate software testers? Ask them how. In: ACM/IEEE International Symposium on Empirical Software Engineering and Measurement Would, pp. 95–104 (2017)
73. Ma, Z., et al.: A data-driven risk measurement model of software developer turnover. Soft. Comput. **24**(2), 825–842 (2019). https://doi.org/10.1007/s00500-019-04540-z
74. Sharp, H., Baddoo, N., Beecham, S., Hall, T., Robinson, H.: Models of motivation in software engineering. Inf. Softw. Technol. **51**, 219–233 (2009). https://doi.org/10.1016/j.infsof.2008.05.009
75. Silva, F.Q.B., Franca, A.C.C.: Towards understanding the underlying structure of motivational factors for software engineers to guide the definition of motivational programs. J. Syst. Softw. **85**, 216–226 (2012). https://doi.org/10.1016/j.jss.2010.12.017
76. Bordel, B., Alcarria, R.: Assessment of human motivation through analysis of physiological and emotional signals in Industry 4.0 scenarios. J. Ambient. Intell. Humaniz. Comput. 1–21 (2017). https://doi.org/10.1007/s12652-017-0664-4
77. Lenberg, P., Feldt, R.: Psychological safety and norm clarity in software engineering teams. In: CHASE Cooperative and Human Aspects of Software Engineering, pp. 79–86 (2018)
78. Felderer, M., Travassos, G.: Contemporary empirical methods in software engineering. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-32489-6
79. Fritz, T., Murphy, G.C., Murphy-Hill, E., Ou, J., Hill, E.: Degree-of-knowledge: modeling a developer's knowledge of code. ACM Trans. Softw. Eng. Methodol. **23**(2), 1–42 (2014)
80. Cárdenas-Castro, C., Gil Julio, J.C., Rodríguez, P.: Soft skills training: performance psychology applied to software development. In: IEEE/ACM 12th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE), pp. 115–116 (2019)
81. Basirati, M.R., Otasevic, M., Rajavi, K., Böhm, M., Krcmar, H.: Understanding the relationship of conflict and success in software development projects. Inf. Softw. Technol. **126**, 106331 (2020). https://doi.org/10.1016/j.infsof.2020.106331
82. Besker, T., Ghanbari, H., Martini, A., Bosch, J.: The influence of Technical Debt on software developer morale. J. Syst. Softw. **167**, 110586 (2020). https://doi.org/10.1016/j.jss.2020.110586
83. Olsson, J., Risfelt, E., Besker, T., Martini, A., Torkar, R.: Measuring affective states from technical debt: a psychoempirical software engineering experiment. Empir. Softw. Eng. **26** (2021)
84. Shastri, Y., Hoda, R., Amor, R.: The role of the project manager in agile software development projects. J. Syst. Softw. **173**, 110871 (2021). https://doi.org/10.1016/j.jss.2020.110871
85. Easterbrook, S., Singer, J., Storey, M.A., Damian, D.: Selecting empirical methods for software engineering research. In: Guide to Advanced Empirical Software Engineering, pp. 285–311 (2008)
86. Kitchenham, B.A., et al.: Preliminary guidelines for empirical research in software engineering. IEEE Trans. Softw. Eng. **28**, 721–734 (2002)
87. Girardi, D., Novielli, N., Fucci, D., Lanubile, F.: Recognizing developers' emotions while programming. In: ICSE International Conference on Software Engineering, pp. 666–677 (2020)
88. de Magalhães, C.V.: Toward understanding work characteristics in software engineering. SIGSOFT Softw. Eng. Notes **41**, 1–6 (2016). https://doi.org/10.1145/3011286.3011

89. Santos, R.E.S., et al.: Work design and job rotation in software engineering: results from an industrial study. In: CHASE Cooperative and Human Aspects of Software Engineering, pp. 139–146 (2019)
90. Graziotin, D., Fagerholm, F., Wang, X., Abrahamsson, P.: On the unhappiness of software developers. In: EASE Evaluation and Assessment in Software Engineering (2017)
91. Graziotin, D., Fagerholm, F., Wang, X., Abrahamsson, P.: What happens when software developers are (un)happy. J. Syst. Softw. **140**, 32–47 (2018). https://doi.org/10.1016/j.jss.2018.02.041
92. Russo, D., Hanel, P.H.P., Altnickel, S., van Berkel, N.: Predictors of well-being and productivity among software professionals during the COVID-19 pandemic – a longitudinal study. Empir. Softw. Eng. **26**(4), 1–63 (2021). https://doi.org/10.1007/s10664-021-09945-9
93. González Llaneza, F.M.: Instrumentos de evaluación psicológica. La Habana (2007)
94. Andersson, D., Rankin, A., Diptee, D.: Approaches to team performance assessment: a comparison of self-assessment reports and behavioral observer scales. Cogn. Technol. Work **19**(2–3), 517–528 (2017). https://doi.org/10.1007/s10111-017-0428-0
95. Passos, C., Cruzes, D.S.: Applying theory of reasoned action in the context of software development practices insights into team intention and behavior. In: EASE Evaluation and Assessment in Software Engineering, pp. 2–11 (2013)
96. Christov, S.C., Hoffman, M.E.: Experiential learning of software project management and software development via course collaboration. In: SIGCSE 2019 - Proceedings of ACM Technical Symposium on Computer Science Education, pp. 2013–2019 (2019)
97. Fritz, T., Begel, A., Müller, S.C., Yigit-Elliott, S., Züger, M.: Using psycho-physiological measures to assess task difficulty in software development categories and subject descriptors. In: ICSCA International Conference on Software and Computer Applications, pp. 402–413 (2014)
98. de Oliveira Neto, F.G., Torkar, R., Feldt, R., Gren, L., Furia, C.A., Huang, Z.: Evolution of statistical analysis in empirical software engineering research: current state and steps forward. J. Syst. Softw. **156**, 246–267 (2019). https://doi.org/10.1016/j.jss.2019.07.002
99. Wackerly, D.D., Mendenhall, W., Scheaffer, R.L.: Estadística Matemática con Aplicaciones (2010)
100. Devore, J.L.: Probability and statistics for engineering and the sciences. Cengage Learning ALL, Canadá (2008)
101. Colomo-Palacios, R., Casado-Lumbreras, C., Soto-Acosta, P., García-Peñalvo, F.J., Tovar-caro, E.: Competence gaps in software personnel: a multi-organizational study. Comput. Human Behav. **29**, 456–461 (2013). https://doi.org/10.1016/j.chb.2012.04.021
102. Marczak, S., Gomes, V.: On the development of a theoretical model of the impact of trust in the performance of distributed software projects. In: 2013 6th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE), pp. 97–100 (2013). https://doi.org/10.1109/CHASE.2013.6614740
103. Bernárdez, B., Cortés, A.R.: A controlled experiment to evaluate the effects of mindfulness in software engineering categories and subject descriptors. In: ESEM Empirical Software Engineering and Measurement (2014)
104. Sutanto, J., Zurich, E.T.H.: Investigating task coordination in globally dispersed teams. ACM Trans. Manag. Inf. Syst. **6**(2), 1–31 (2015)
105. Lenberg, P., Wallgren Tengberg, L.G., Feldt, R.: An initial analysis of software engineers' attitudes towards organizational change. Empir. Softw. Eng. **22**(4), 2179–2205 (2016). https://doi.org/10.1007/s10664-016-9482-0
106. Kropp, M., Meier, A., Anslow, C., Biddle, R.: Satisfaction and its correlates in agile software development. J. Syst. Softw. **164**, 110544 (2020). https://doi.org/10.1016/j.jss.2020.110544
107. Faraway, J.: Linear Models with R. Taylor & Francis (2014)

108. Gren, L.: Standards of validity and the validity of standards in behavioral software engineering research: the perspective of psychological test theory. In: International Symposium on Empirical Software Engineering and Measurement (2018)
109. Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., Tomás-Marco, I.: El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. An. Psicol. **30**, 1151–1169 (2014). https://doi.org/10.6018/analesps.30.3.199361
110. Russo, D., Stol, K.-J.: PLS-SEM for software engineering research. ACM Comput. Surv. **54**, 1–38 (2021). https://doi.org/10.1145/3447580
111. Buse, R.P.L., Zimmermann, T.: Analytics for software development. Proc. FSE/SDP Work. Futur. Softw. Eng. Res. FoSER **2010**, 77–80 (2010). https://doi.org/10.1145/1882362.1882379
112. Adams, W.: Conducting semi-structured interviews. In: Handbook of Practical Program Evaluation, pp. 492–505. Jossey-Bass (2015)
113. Silva, F.Q.B.: The innovative behaviour of software engineers: findings from a pilot case study. In: ESEM Empirical Software Engineering and Measurement (2016)
114. Santos, R.E.S., da Silva, F.Q.B., de Magalhães, C.V.C., Monteiro, C.V.F.: Building a theory of job rotation in software engineering from an instrumental case study. In: 2016 IEEE/ACM 38th IEEE International Conference on Software Engineering Building, pp. 971–981 (2016)
115. Minetto, B., et al.: Synthesizing researches on knowledge management and agile software development using the meta-ethnography method. J. Syst. Softw. **178**, 110973 (2021). https://doi.org/10.1016/j.jss.2021.110973
116. Ralph, P., Kelly, P.: The dimensions of software engineering success. In: ICSE International Conference on Software Engineering, pp. 24–35 (2014)

Effects of Debriefing in Computer-Supported Collaborative Learning Pyramid Scripts with Open-Ended Task

Valguima Odakura[1](✉), Ishari Amarasinghe[2], Davinia Hernández-Leo[2], Roberto Sánchez-Reina[2], Emily Theophilou[2], and René Lobo-Quintero[2]

[1] Universidade Federal da Grande Dourados (UFGD), Dourados, Brazil
valguima.odakura@gmail.com

[2] ICT Department, Universitat Pompeu Fabra, Barcelona, Spain
{ishari.amarasinghe,davinia.hernandez-leo,roberto.sanchez,
emily.theophilou,renealejandro.lobo}@upf.edu

Abstract. In this study we investigate the benefits that debriefing can add to collaborative Pyramid script with open-ended tasks. The open-ended task allows students to produce multiple possible solutions to a given problem and requires learners to express personal opinions based on previous experiences and intuitions. In this sense, misconceptions and gaps can appear in the collaboration process, demanding a teacher intervention. Debriefing, as part of teacher orchestration tasks, enables teachers to facilitate students' reflection about the learning experience, correcting mistakes and filling gaps. Qualitative analysis of students' answers through concepts and their relations was developed. The examination of concepts and relations supported the benefits of the Pyramid script with open-ended tasks and debriefing to learning. The results of the study indicate that students increased the concepts mentioned and built more relations between concepts after debriefinLg.

Keywords: Computer-supported collaborative learning · Scripts · Pyramid script · Open-ended task · Debriefing

1 Introduction

Computer-Supported Collaborative Learning (CSCL) scripts structure collaboration interactions in order to facilitate learning. CSCL scripts are important when free collaboration does not result in interaction and consequently in learning [3]. CSCL scripts structure the process of interactions, defining sequences of activities, distributions of groups, roles and resources [3]. An implementation of CSCL script based on Pyramid collaborative learning flow pattern is the PyramidApp. A Pyramid flow is initiated with students solving a task individually. Then, in a second level of the Pyramid, in small groups, the individual solutions

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 23–37, 2022.
https://doi.org/10.1007/978-3-031-20218-6_2

are discussed and refined into a common answer. In the next levels of the Pyramid, larger groups are iteratively formed from small groups and group discussions continue refining the previous solution to a commonly agreed solution [10].

Pyramid activity is being used in collaborative learning activities with observable impact in learning gains [2]. Some factors that could influence the design of Pyramid activities are: the pedagogical envelope, the type of tasks, pyramid design elements, and the need for epistemic orchestration and debriefing [1].

Collaborative learning tasks can be open-ended or closed-ended. Closed-ended tasks have one correct answer that can be “yes” or “no” answer or a limited set of possible answers. In open-ended tasks, students can follow multiple solution paths to arrive at or to produce multiple possible solutions and elaborations to a given problem and often require learners to make judgments and express personal opinions or beliefs [12].

Prior research has provided first insights that pyramid activities can increase students’ learning gains, measured in terms of an increased level of precision and a decreased level of confusion associated with an answer. However, in some cases learning gains immediately after participating in PyramidApp activity do not seem to significantly improve in terms of precision and confusion, especially if we consider the type of task. In [1] it was reported 3 different learning activities, 2 with closed-ended tasks and 1 with the open-ended task. The 2 activities with closed-ended tasks presented learning gains in terms of increased precision and decreased confusion, however, the open-ended one did not lead to learning gains. From another work, [2] it was presented 4 learning activities, 2 with closed-ended tasks and 2 with open-ended ones. The results were learning gains in terms of increased precision and decreased confusion for closed-ended tasks. However, the learning gain was not observed in the open-ended tasks. Moreover, both studies noticed decreased precision and increased confusion after Pyramid activity.

In [2] it was added to the learning activities a debriefing phase after the Pyramid. Notably, the learning gain appeared in terms of increased precision and decreased confusion after the teacher-led debriefing for both types of tasks. For the open-ended tasks, in one case the learning gain outperformed the individual answer and for the other case, it only outperformed the post Pyramid learning gain.

Debriefing activities require the teacher to elaborate on students’ responses in real-time, being a demanding task in terms of orchestration. Orchestration refers to the real-time management of learning scenarios by the teacher [4].

In this study we are interested in type of tasks and debriefing factors, once prior research indicates that Pyramid activity with open-ended tasks followed by debriefing impacts learning gains. Moreover, how debriefing after Pyramid activity with open-ended task influences learning outcomes is not fully known. More research is needed to provide evidence of the benefits that debriefing can add to scripted collaboration with open-ended tasks. To this end, the research question proposed in this study is: “How do open-ended tasks affect collaborative experience with Pyramid activity and post debriefing?”

This paper is organized as follows. In Sect. 2 details about debriefing in collaborative learning activities are presented. In Sect. 3 the methods followed are explained. In Sect. 4 the results of the study are presented, followed by Sect. 5 which provides a discussion of the study findings. Finally, in Sect. 6 is provided concluding remarks and future research directions.

2 Debriefing and Collaborative Learning Activities

The term debriefing is used in different domains, as military training and psychological approach, enabling participants to review the facts and thoughts after an event. In educational settings or experience-based learning, debriefing is used as post-experience analysis, in simulations and game-based learning tasks [9].

Debriefing is a form of reflective practice and provides a means of reflection-on-action in the process of continuous learning. The idea behind debriefing is the belief that experience alone does not lead to learning, but rather the deliberate reflection on that experience [13]. However, reflection after a learning experience might not occur naturally, or if it does, it is unsystematic [6]. In this manner, conducting a formal debriefing focused on the reflective process is used as part of the learning process [6,13].

The debriefing can occur after the experience, the post-event debriefing, or during the event, the within-event debriefing, through interruptions to students' actions when mistakes occur [13]. The post-event-debriefing can be facilitator-guided or self-guided, when performed by individuals or conducted by teams [13]. The teacher facilitated post-event-debriefing is the recommended and most widely practiced method [5].

The design of the debriefing session must be adapted to the learning objectives and characteristics of the participants [6]. Seven common structural elements involved in the debriefing process were proposed by [9]: debriefer; participants to debrief; an experience; the impact of the experience; recollection; report and time.

In the context of collaborative learning, the experience to be reflected is the activity performed in CSCL script. In this way, ArgueGraph script activates argumentation among members and closes the activity with debriefing, where the teacher organizes the arguments produced by students, articulating them with theories [8]. In the Concept Grid script, individual students work with a part of knowledge and groups are formed composed of students with different parts of knowledge, who collectively solve a problem that requires knowledge of each of them. In the debriefing session, the teacher compares the solutions produced by different groups and requests them to explain the distinctions [3].

In this study collaborative learning is addressed by considering the learning design, processes, and outcomes. The learning design refers to group formation, type of task and type of education. The groups are formed randomly by the PyramidApp, the first group level with 3–4 students and the next group level with 7–11 students. The type of task is open-ended and the type of education is informal, conducted in a workshop format. The processes of collaboration refer to collaborative Pyramid activity, with individual and group phases followed by teacher-led debriefing. Finally, the outcomes of collaboration refer to individual achievements in terms of concepts and their relations. Figure 1 illustrates this framework for investigation of collaborative learning.

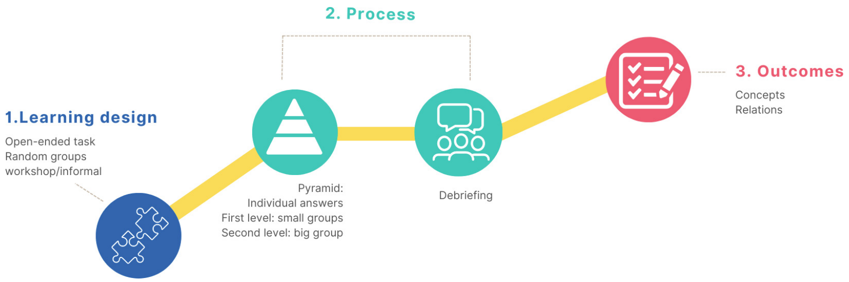


Fig. 1. Investigating collaborative learning: learning design - process - outcomes.

The experience or practice to be reflected in the context of collaborative learning is the activity performed in CSCL script. Besides the PyramidApp dashboard is not designed to support debriefing, the dashboard orients the teacher to conduct the debriefing at the end of the activity based on the winning answers from students' groups.

3 Methods

3.1 Participants and Context

A quasi-experimental study was conducted in a public high school in Brazil. The sample of the study consisted of 33 students distributed in 4 groups of 7–11 students from 2nd and 3rd year, aged from 16 to 18 years old. Data were collected in the context of media literacy workshops conducted by one teacher. Students provided their informed consent for data collection.

3.2 Tools

Each group participated in the same collaborative learning activity with the same open-ended task. Pyramid activity consisted of three levels, an individual submission level, and two group levels. The Pyramid activity duration was designed for 16 min and had a slight difference from one group to another, based on students' needs and teacher orchestration, i.e., in some cases the teacher added 1 more minute to allow students to finish a level. The open-ended task, enabling students to make judgments and express personal opinions or beliefs, was about social media awareness, asking students: “How do you think social media influences our body and appearance? That is, how we feel about our body image.”. Table 1 presents the details of activity.

Table 1. Collaborative learning activity details.

Session	Number of students	Duration	Open-ended task given to students
A	7	16 min	How do you think social media influences our body and appearance? That is, how we feel about our body image
B	7		
C	8		
D	11		

The PyramidApp tool provides an authoring space where teachers can design CSCL activities. A screenshot of the design space of the PyramidApp is shown in Fig. 2, where it is possible to configure several parameters as the number of students expected, the number of levels in the Pyramid script, the number of students per level and the duration allocated to different phases.

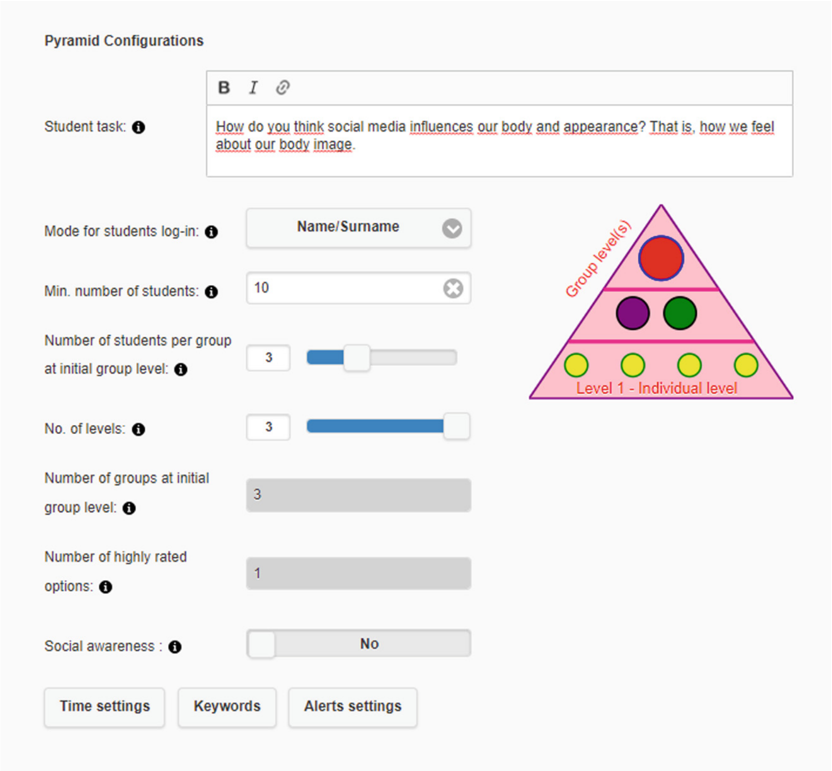
PyramidApp collaboration structure follows some levels. Students give an individual answer for a given problem (option submission level). Then small groups of students are randomly formed (first group level). In the small groups, students examine the answers submitted by other students individually (rating). Students then take part in a discussion at the small group level and improve existing options collaboratively (improving). Larger groups are formed automatically by merging small groups (second group level). In the larger groups, students participate in an individual rating of the answers selected from the previous level (rating) and then collaboratively improve the selected answers (improving). Teacher can orchestrate collaboration at all levels through the PyramidApp dashboard. A screenshot of the PyramidApp dashboard is presented in Fig. 3.

The students were given training on how to use the PyramidApp for collaboration prior to the experimental session reported in this study.

3.3 Debriefing in Pyramid Scripts

After the Pyramid activity a debriefing was conducted by the teacher based on answers produced during the Pyramid activity and adding concepts for filling gaps if needed. The students' intervention during the debriefing guided the direction of the discussions. The debriefing lasted around 15 min.

Following the seven common structural elements involved in the debriefing process [9], the debriefer is the teacher that will conduct the debriefing and the participants are the students. The experience is the collaborative activity in the PyramidApp and the impact of the experience depends on the relevance of the experience for the students and should clarify the facts, concepts, and principles. The students participate with the recollection of their experience and report the experience in a verbal manner. The time is post-event debriefing.



Pyramid Configurations

Student task: How do you think social media influences our body and appearance? That is, how we feel about our body image.

Mode for students log-in: Name/Surname

Min. number of students: 10

Number of students per group at initial group level: 3

No. of levels: 3

Number of groups at initial group level: 3

Number of highly rated options: 1

Social awareness : No

Fig. 2. PyramidApp user interface.

The main objective of debriefing is to promote a reflective process. For this, the teacher used the same strategy for all groups. The debriefing starts from the group answer provided by students in the Pyramid dashboard, i.e., the winning answer developed for the group. If this answer has some gaps, the missing concepts are added by the teacher. But, different from feedback, that is one way intervention, the debriefing process allows interaction and reflective discussions. For this, the teacher asks some questions to students, addressing missing or confusing topics. The students can explain their ideas and the teacher can guide them to reflection. The conversation evolves and confusion and doubts are solved by the teacher as it appears. Finally, the teacher summarized the conversation, highlighting the points discussed for the group.

Following debriefing, the students were asked to respond to a final questionnaire which asked them to write an answer to the same Pyramid task.

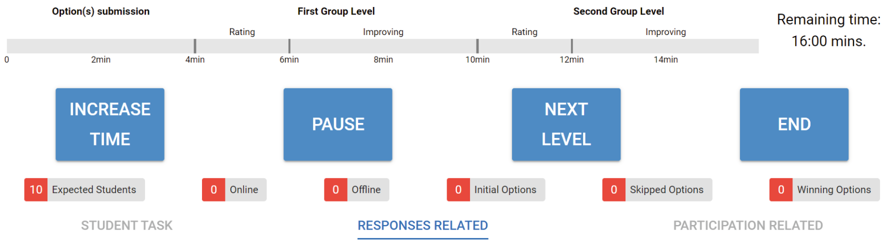


Fig. 3. PyramidApp dashboard: option submission, first group level and second group level.

3.4 Data Collection and Measurement

In this study the data was collected in three moments: an individual answer at first level of Pyramid, an individual answer after Pyramid and an individual answer after debriefing. Figure 4 summarizes the data collection.

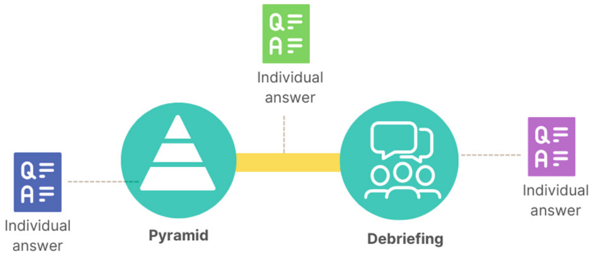


Fig. 4. Data collection.

In order to analyze the outcomes of collaboration, the student's learning gains can be measured in terms of an increased level of precision and a decreased level of confusion associated with an answer. Levels of precision can range from 0 (not precise) to 3 (student's response matches teacher's response). Levels of confusion can range from 0 (None) to 3 (high) [1].

4 Results

The students' answers were evaluated by the teacher grading the levels of precision and confusion. In Fig. 5 it is possible to see the learning gains in terms of average precision for pre, middle and post answers, which refers to the following, respectively, individual answers submitted to Pyramid activity, answers after the collaborative Pyramid activity and after debriefing. The results from Fig. 5 are similar to [2] in terms of decreased precision after Pyramid activity and increased precision after debriefing. In these sessions, there was no confusion in the answers.

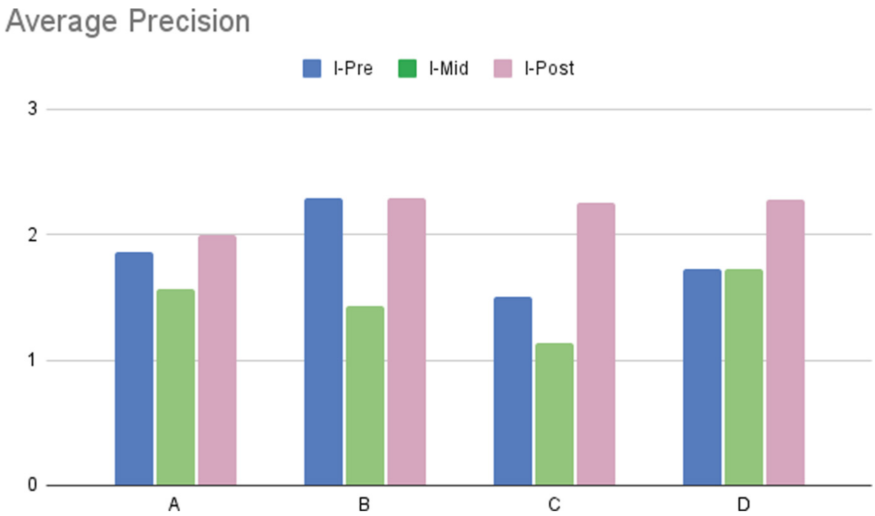


Fig. 5. Average precision in students' answers from A, B, C and D in I-Pre (individual Pyramid answer), I-Mid (individual, after Pyramid task) and I-Post (individual, after debriefing).

As shown in Fig. 5 in all sessions the precision of individual students' answers have decreased after participating in the Pyramid activity. For all sessions, the precision of individual students' answers has increased after debriefing. For sessions A, C and D the precision increased from the pre to the post, i.e., the students presented learning gain in terms of precision. Despite that, in session B the precision of students' answers increased after debriefing, but it was the same as the prior answers, i.e., the students did not present learning gains from the initial to the end of the session.

In order to illustrate the differences from pre to post, we can look at the answer from one student: Pre: *Can influence positively or negatively, according to the profiles we follow.* Mid: *It influences a lot.* Post: *Social media can affect positively or negatively, as they require many beauty and body standards, standing out as "healthy" standards of living.* In this example it is possible to note that from pre to post the student added concepts and explanations in his/her answer.

A qualitative analysis of the answers presented by the students was carried out to better understand the results of Fig. 5. The answers were coded considering the concepts treated by students. We followed an inductive data coding approach [14]. Firstly, the initial codes are extracted based on a preliminary read of the answers. In a second reading, the codes are refined and the answers are marked with the codes found. Following that, a new reading of the answers was conducted to review the codes extracted from the answers. The final coding scheme consisted of the following codes: comparison, pattern, false, unreal, negative and adapt. The codes, their meanings and an example in the answer are presented in Table 2.

Table 2. Coding scheme.

Code	Meaning	Context
Comparison	Student mentions that he/she makes comparisons when looking at body images in social media	*"you can see body images and compare yourself"*
Pattern	Student is aware of the idealized body images presented as patterns in social media	*"social media demand many standards of beauty and body,"*
False	Student is aware of false and manipulated images in social media	*"(body images) has several changes with effects and edits"*
Unreal	Student is aware that body images shown on social media are unreal and unattainable bodies	*"on social media, many people have the image of a perfect body, something that does not exist"*
Negative	Student reports negative emotions like envy, depression or sadness	*"can make you feel extremely bad about yourself", "putting us down", "we envy"*
Adapt	Student declares that he/she needs to change his/her body to adapt to the body pattern presented in social media	*"wanting to follow a pattern that are posted by influencers on the social media"*

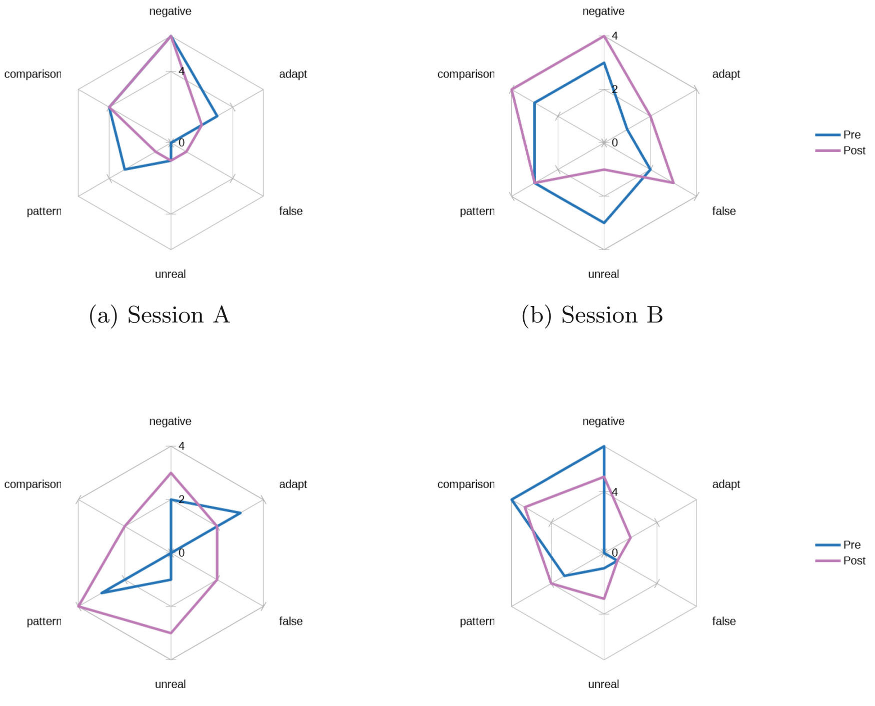
Figure 6 illustrates the frequency of codes appearing in students' individual Pyramid answers (pre) and after debriefing answers (post) for the 4 sessions for the codes comparison, pattern, false, unreal, negative and adapt.

For session A it is possible to note that students mentioned false/manipulated images more times after debriefing. However, comparison and negative emotions remained in the same frequency, but idealized body image patterns and the need to change to adapt to the pattern decreased.

For session B it is possible to note that students mentioned comparison, negative emotions and false/manipulated images more times after debriefing. However, the need to change to adapt to the pattern remained in the same frequency, but unreal/unattainable body image decreased.

Session C presented the most difference from pre to post. It is possible to note that students mentioned four of five terms more times after debriefing.

In session D it is possible to note that students mentioned the need to change to adapt to the pattern, unreal/unattainable body image and false/manipulated images more times after debriefing. However, false/manipulated images remained in the same frequency, but comparison and negative emotions decreased.



(a) Session A

(b) Session B

(c) Session C

(d) Session D

Fig. 6. Chart code frequencies for sessions A, B, C and D.

From all sessions, the knowledge of the topic was explored by the students and it was strengthened after debriefing.

Besides code frequency, we also explored the relations among concepts presented in students' answers. A network graph allows analyzing elements that stand in pairwise relations [11]. The relations may be qualitative, present or not; they may also be directed (from one element to another, but not the other way); and they may also be quantitative, i.e., they may possess weights. The network analysis begins by defining a specific corpus of texts, in this study, the students' answers for the task. From the students' answers, it was extracted the codes as shown in Table 2. The codes are the elements of the network. The relations between elements are extracted from the answers. If a student reports that "*I compare myself with edited body image in social media and it makes me feel depressed.*", this answer has 3 codes: false, comparison and negative. For this answer the codes false, comparison and negative are related, representing 3 edges at the network: false with comparison, false with negative and negative with comparison. The network graphs for the 4 sessions are presented in Fig. 7.

In Fig. 7 it is possible to see the relations among codes for the 4 sessions in two moments: pre, before group Pyramid activity and post, after debriefing. In session A it was added the code for false images from pre to post and it changed the relations they did. For sessions B, C and D it is possible to note that students were able to relate more codes at the end, visually represented by more links between nodes. And, the stronger the link, the more times the same relation was found in students' answers.

The students reported the relationships among idealized body image patterns presented in social media images, which are false/manipulated images, showing unreal and unattainable bodies, when compared to real bodies, can lead to negative emotions such as envy, anxiety, and depression.

The relations between concepts can be represented by the degree of a node, i.e., the number of its neighbors [11]. For example, in Fig. 7, session D - pre, the code comparison has 4 relations and the code false has only one. Considering we have 6 codes, a fully connected network or a complete network would have all nodes with 5 relations each. In this manner, if students articulated all concepts in the same answer it would result in a complete network. From Table 3 it is possible to see the sum of the degrees of all nodes for the 4 sessions. If we consider that the sum of the degrees of all nodes for a complete network is 30, we can see that sessions B, C and D enriched their connections after debriefing, approximating from a complete network. Session A did not increase the number of relations, however, it increased the concepts considered, that was 5 before the activity and 6 in the post debriefing.

Table 3. Sum of the degrees of all nodes in the networks from Fig. 7.

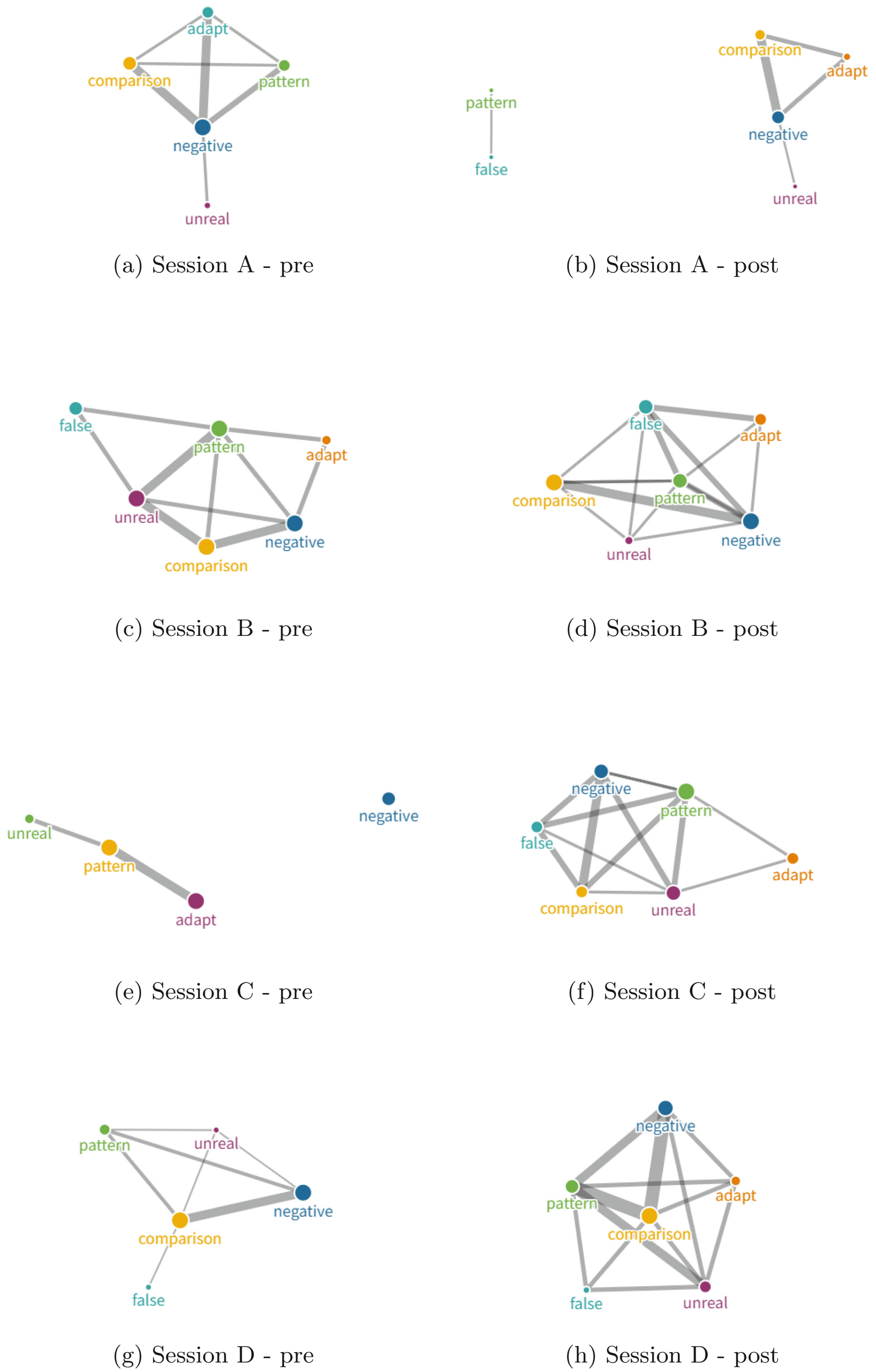
Sessions	Pre	Post
A	13	10
B	20	26
C	12	24
D	14	26

5 Discussion

From the learning activity conducted, the open-ended task resulted in similar levels of learning gains for the 4 sessions studied, which corroborate with previous works [1,2]. The results are decreased precision after Pyramid activity and increased precision after debriefing, as shown in Fig. 5.

We can not observe explicit (individual) learning gains when comparing individuals' answers prior to the Pyramid with individuals' answers after Pyramid. In a detailed look at the answers, we could note that most of the answers after Pyramid are incomplete, with only one word or small phrases. The answers after debriefing are complete answers, allowing students to explain their knowledge. To this extent, we argue that students do not decrease knowledge after Pyramid activity.

From the learning design (open-ended task and workshop session) we derived three assumptions about the process (Pyramid and debriefing) and outcomes



(a) Session A - pre

(b) Session A - post

(c) Session B - pre

(d) Session B - post

(e) Session C - pre

(f) Session C - post

(g) Session D - pre

(h) Session D - post

Fig. 7. Network graphs from sessions A, B, C and D in pre: individual Pyramid answer and post: after debriefing.

(students' answers): (1) students are less inclined to give complete answers after Pyramid, so we indicate a future motivation inspection at Pyramid activity; (2) the open-ended task does not have a right/wrong answer, so students can provide incomplete answers that could be partially correct. It is more difficult to occur at a closed-ended task that has a correct answer; (3) the intervention was conducted in a workshop and was an informal, non-curricular activity so that students were free to participate or not.

During a Pyramid activity based on an open-ended task, students have the opportunity to discuss multiple possible solutions for the task and can express personal opinions or beliefs. This type of task is aligned with Pyramid activity as it allows developing knowledge collaboratively, considering the individual initial beliefs and constructing agreements for a refined solution. On the other hand, this type of question can propagate confusion and some students can not be persuaded by the group, once they can remain with their prior beliefs, even if it is a misconception. Otherwise, the students can work on partial solutions, missing some concepts or relations between concepts. In these particular cases, debriefing is a highly recommended practice after the Pyramid activity to fill gaps and correct misconceptions. At the same time debriefing contributes to deriving useful insights through a discussion of the experience.

Returning to the question: How do open-ended tasks affect collaborative experience with Pyramid activity and post debriefing?

Our qualitative analysis of concepts and relations in the students' answers to an open-ended task confirm that the process of Pyramid activity followed by a teacher-led debriefing impacts the learning outcomes. The concepts represented by the codes extracted from the students' answers reveal that students expanded the concepts they mention from individual Pyramid answers to answers after debriefing (Fig. 6). More than that, students increased the relation of concepts from prior to post, making more relations between concepts (Fig. 7).

From social media workshop perspective, students could articulate the main concepts related to social media body image, referring to idealized body image patterns presented in social media images, which are mostly false or manipulated images, showing unreal and unattainable bodies, when compared to real ones can lead to negative emotions such as envy, anxiety, and depression. These answers are strongly connected with the studies in the area [7,15].

Considering an open-ended task, the debriefing can take different ways, depending on the students' contributions. The conversation is always enriched by the students' personal opinions or beliefs. An important part of debriefing is to guarantee a safe psychological place for students to share their beliefs, as stated by [13]. That is crucial for a successful debriefing because if students do not talk about their misconceptions, the teacher is not able to discuss them and promote a productive reflection.

Finally, the debriefing was conducted for the same teacher at all sessions. It is worth noting that the debriefer experience can impact the debriefing results. In this way, having a teacher use a debriefing script may improve the ability of facilitators to effectively lead the debriefing conversation, as suggested by [13].

6 Conclusions and Future Work

In this study we investigate collaborative learning in the case of a Pyramid CSCL script addressing an open-ended task and teacher-led debriefing. Prior preliminary research indicates that a Pyramid activity with open-ended tasks followed by debriefing impacts learning gains. This work contributes with additional evidence that corroborates and extends this previous work. The novelty is related to the context and learning design used for the data collection as well as to the methodology used to analyze the learning outcomes. The context is an informal learning setting involving teenage students, aged 16 and 18. In terms of the learning design, the task is open-ended and of a nature that leads students to express personal opinions based on previous experiences and institutions. Regarding the methodology, we have analyzed the evolution of concepts and relations in the evolution of students' expressed knowledge from a perspective of the learning outcomes.

Study findings support the importance of debriefing to learning gains achieved during scripted CSCL activities with open-ended tasks as it summarizes learning experience, fills gaps and corrects mistakes. The students were able to state more concepts and articulate them in more relations after the debriefing. Promising results about debriefing encourage future work on how debriefing influences teachers' orchestration load.

There are several limitations to this study. The number of cases we considered is low and the number of students participating in each activity is relatively low. These limitations can have an impact on the obtained results. However, as exploratory research there is no attempt to generalize the findings to a wider population, but to gain insights into collaborative learning. Another limitation is that only one teacher participates in the sessions, which means that the debriefing could have other results with a different teacher. For future work we plan to investigate the debriefing scripts that could guide teachers to structured debriefing. It is also relevant to explore debriefing time, i.e., the differences between post-event debriefing and within-event debriefing. Considering the cases in which confusion is propagated in the group, an earlier intervention by the teacher, within-event debriefing, could address mistakes as they appear and improve learning.

Acknowledgements. This work has been partially supported by the grants PID2020-112584RB-C33, MDM-2015-0502 funded by MICIN/AEI/10.13039/501100011033 and the Volkswagen Stiftung (Courage, ref. 95 566). D. Hernández-Leo (Serra Húnter) acknowledges the support by ICREA under the ICREA Academia programme.

References

1. Amarasinghe, I., Hernández-Leo, D., Theophilou, E., Roberto Sánchez Reina, J., Quintero, R.A.L.: Learning gains in pyramid computer-supported collaboration scripts: factors and implications for design. In: Hernández-Leo, D., Hishiyama, R., Zurita, G., Weyers, B., Nolte, A., Ogata, H. (eds.) CollabTech 2021. LNCS,

vol. 12856, pp. 35–50. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-85071-5_3

2. Amarasinghe, I., Hernández-Leo, D., Manathunga, K., Pérez, J.C., Dimitriadis, Y.: Teacher-led debriefing in computer-supported collaborative learning pyramid scripts. In: International Conference on Computer-Supported Collaborative Learning (CSCL). pp. 171–178. International Society of the Learning Sciences (2022)
3. Dillenbourg, P.: Over-scripting CSCL: the risks of blending collaborative learning with instructional design. Three worlds of CSCL. Can we support CSCL, vol. 61491 (2002)
4. Dillenbourg, P.: Design for classroom orchestration. Comput. Educ. **69**, 485–492 (2013)
5. Dufrene, C., Young, A.: Successful debriefing-best methods to achieve positive learning outcomes: a literature review. Nurse Educ. Today **34**(3), 372–376 (2014)
6. Fanning, R.M., Gaba, D.M.: The role of debriefing in simulation-based learning. Simul. Healthc. **2**(2), 115–125 (2007)
7. Franchina, V., Coco, G.L.: The influence of social media use on body image concerns. Int. J. Psychoanal. Educ. **10**(1), 5–14 (2018)
8. Jermann, P., Dillenbourg, P.: Elaborating new arguments through a CSCL script. In: Andriessen, J., Baker, M., Suthers, D. (eds.) Arguing to Learn, vol. 1, pp. 205–226. Springer, Dordrecht (2003). https://doi.org/10.1007/978-94-017-0781-7_8
9. Lederman, L.C.: Debriefing: toward a systematic assessment of theory and practice. Simul. Gaming **23**(2), 145–160 (1992)
10. Manathunga, K., Hernández-Leo, D.: Authoring and enactment of mobile pyramid-based collaborative learning activities. Br. J. Edu. Technol. **49**(2), 262–275 (2018)
11. Painter, D.T., Daniels, B.C., Jost, J.: Network analysis for the digital humanities: principles, problems, extensions. Isis **110**(3), 538–554 (2019)
12. Reja, U., Manfreda, K.L., Hlebec, V., Vehovar, V.: Open-ended vs. close-ended questions in web questionnaires. Dev. Appl. Stat. **19**(1), 159–177 (2003)
13. Sawyer, T., Eppich, W., Brett-Fleegler, M., Grant, V., Cheng, A.: More than one way to debrief: a critical review of healthcare simulation debriefing methods. Simul. Healthc. **11**(3), 209–217 (2016)
14. Thomas, D.R.: A general inductive approach for analyzing qualitative evaluation data. Am. J. Eval. **27**(2), 237–246 (2006)
15. Tiggemann, M., Anderberg, I.: Social media is not real: The effect of ‘instagram vs reality’images on women’s social comparison and body image. New Media Soc. **22**(12), 2183–2199 (2020)

The Effect of Pair Programming on Code Maintainability

Mamoun Nawahdah[1](✉) and Majdi Jaradat[2]

[1] Birzeit University, Birzeit, Palestine
mnawahdah@birzeit.edu
[2] Ministry of Public Works and Housing, Jerusalem, Palestine

Abstract. Software maintainability is an important key aspect in software development life cycle. It has a huge impact on the time, effort, and cost. This research investigated the effect of applying pair programming on software maintainability vs individual programming approach. We reviewed many related studies which talked about the two approaches, but few of them focused on maintainability. Therefore, we conducted an experiment in educational environment. The participants were novice students in their second year, who finished object oriented programming course. The participating students were divided into two groups, some worked as individual and the others worked as pairs. In order to get good results, we made the experiment among the students in a competitive way. Enabling those who produced the best code to win a cash prize. Finally, code obtained from both approaches was analyzed by software quality assurance expert, who has finished a master degree in software engineering. The results taken from pair groups showed a statistically significant difference in building software with 50% less errors, 30% less code, and 25% better code. Furthermore, there is a marginal reducing in programming time particularly in small size software compared to individual programming.

Keywords: Pair programming · Solo programming · Individual programming · Maintainability · High quality code

1 Introduction

Pair-Programming is one of the most important practices of Extreme Programming (XP) under agile technology. The concept of pair-programming is that two programmers sit next to each other, work together to produce the same code [1,4–7]. They use the same computer, one keyboard, one mouse and sharing the same screen. One programmer considered as driver his responsibility to write code and the other as navigator his responsibility to revise written code. Furthermore, navigator suggesting improvements on the written code, in addition to code inspection task. The driver sit to the right, while the navigator sit to the left as shown in Fig. 1, they taking turns to code typing. An idiom says "two

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 38–51, 2022.
https://doi.org/10.1007/978-3-031-20218-6_3

heads are better than one" referring to the advantages of collaborative work. The role played by both driver and navigator is pivotal, in order to create high quality code. So, producing robust software, sustainable, maintainable and cost effectiveness, related to high quality code [8].

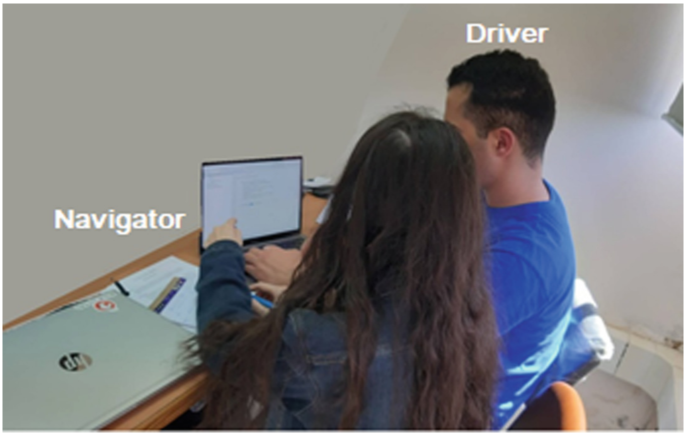


Fig. 1. Two programmers practicing pair-programming in the lab

Pair programming concepts are not recent, but it recently attracted more attention in software development, especially on software maintainability factor. Many research's showed that pair programming decreases the number of bugs [3–5]. Other research revealed that pair-programming produces understandable base code, eliminate software complexity, and build high quality code [1,2,5]. The time is very important to make the software ready to be deployed, so, pair programming has proven to be effective in accelerating productivity, and also to be able to adapt environment changed with minimal risk [3,4].

Pair programming technique comes from the immense popularity it has gained over recent years for its capability of writing high-quality code. For that, software professionals adopt pair programming in every possible situation. They applied the standards and quality constraints in developing [3]. By reviewed related studies, many similar studies have been carried out in educational sector. But, few of them investigated the effect of pair programming on software maintainability.

Maintainability considered as an important issue to focus on. The adoption of applying digital transformation in the world is increasing significantly. So, studying the impact of pair programming on software development in term of maintainability is either increased. Many factors are closely related to achieving maintainability, such as number of errors, number of lines of code, and high code quality. These factors have a big role in building maintainable, understandable and reusable software [1,5,9,12].

In this study, in order to investigate the effect of pair-programming in software maintainability, we conducted an experiment which it was applied on 15 students. The students asked to build a binary calculator program. We divided them into two groups, pair groups and individual groups. The groups adhered to specific criteria to carry out the task. The produced code from each group has been analyzed by software quality assurance (SQA) expert. The results revealed that pair-programing groups produced code 50% less errors, 30% less code and 25% better code.

2 Literature Review

Several experiments have been conducted on pair- programming technique, in both educational and industrial environment. The reviewed studies in this research, covered a common programming problems, and different aspects tackle by pair-programming.

2.1 Pair-Programming

Pair programming consists of a driver and a navigator, sit next to each other. They work together at the same computer, collaborating on the same code [1,4], using one keyboard, one mouse, and sharing the same screen. Navigator guides the driver to right direction [6]. Pair groups can be formed either randomly or by participants who select their prefer partners. Some studies have shown that random pairs spent the lowest time of 16.57 min on the average to find and fix the bug. Junior individual programmers spent 21.88 min/bug [1]. Interaction and communication between teams is very important in software development. It helps building a professional software developers. In educational environment, pair programming reduces the rate of absence. Furthermore, encourages students to complete course requirements required to pass the course. [4]. Students satisfaction in pair- programming condition was better, fairly successful in pedagogy [4,7,10]. Pair programming has performed better than individual programming, and tackles programming tasks in different ways. Pair programming become an effective strategy for learning programming over individual programming technique. Students produce higher quality code using pair programming. In addition to that, they display increased confidence and gain better grade on exams [3,4]. Pairing between team members must be right to generate desirable software. Otherwise, they may be gained a negative impact in terms of lose interest and focus [5]. Partner pairing selection is critical for successful pair programming. For knowledge sharing and educational purpose, we should use novice - expert. But in case of increasing collaboration among group members, two members should be novice-novice. In case of industrial environment, in order to produce high quality code and sustainable, the group members must be from expert [13]. Building an expert team is mandatory in industrial environment, and a key aspect to be competing among markets [8].

2.2 Maintainability

Maintainability is defined as the degree to which an application is understood, repaired, or enhanced [1]. Software maintainability is important because it is approximately 75% of the cost related to any project. That's mean; every one dollar paid on development, needs two or three dollars on maintenance. Maintainability is a software attribute, which plays an important role to predict whether the software with high quality and understandable [5]. Maintainability is a compound component of many attributes, like, understandability, testability, modifiability, reusability [9].

Pair-programming has the ability to correct errors with lowest time nearly 25% on the average compared to individual [1]. Pair-programming reducing the defect ratios, produced less code and better design [5]. Moreover, they decrease programming time, but very marginal, especially in small projects [2]. Pair-programming has performed better than individual programming in term of high-quality code [3,4].

The score of finding bugs by pair programming team in the application was 47%, while in the individual groups was 31% [6]. Moreover, test cases written by students who applied pair programming techniques increase the number of killed mutants with better code coverage.

Reusing what others have already tackled and created can save cost and time, increase productivity, and minimize bugs. So, it's important to have a good expertise, and it should be easy to find among team members. To deepen expertise as effectively, it's required to sharing it, which it can be achieved by pair-programming [9]. Pair programming for teaching pair testing and pair design has achieved an influence in productivity and a higher efficiency. It reduces defect rate by 50%, writing shorter code and improve code readability [11].

3 Research Hypothesis and Methodology

In this research we use the experimental research to manipulate independent variables in order to determine the causes of dependent variables. We started with identified three null hypotheses to be tested to find out if there is difference between using individual programming or pair programming on maintainability factors.

3.1 Null Hypotheses

Our null hypothesis supposes that when we adopt pair programming rather than individual programming, there is no effect on the following factors:

- There is no significant difference in reducing number of errors.
- There is no significant difference in reducing number of code lines.
- There is no significant difference in writing high quality code.
- There is no significant difference in reducing programming time.

Thus, there is no effect on code maintainability. The independent variable in our experiment is the programming settings which has two conditions pair programming and individual programming. In the other hand, the dependent variables are the factors in this experiment which we want to investigate and measure whether pair programming has effect on it or not. Which are:

- Number of errors.
- Number of line of code.
- Code quality.
- Programming time.

3.2 Experiment Scope

We conducted an experiment in educational domain executed in local university.

3.3 Experiment Components

Subjects. The experiment were executed among 15 out of 180 computer science students. They were finished an object oriented programming course in Java language, and they were at the same level of knowledge. We also to consider that all selected students who were participating in the experiment have no pair programming experience before, to get more precise results.

Assignment Method. Group distribution was according participants preferences. Each participant chose to work individually or with a partner. The participants who chose to work pairs, selected their partners based on their desired. The 15 participants were divided into two groups, 10 in pair programming and 5 in individual programming. Each group just works with one condition, so, the experiment design was between group to achieve the following advantages:

- To guarantee that there isn't learning effect from another condition.
- It consumes a shorter time in executing the task.
- Participants do not feel frustrated and tired of carrying out the task.

The only disadvantage of between group is that we cannot determine the personal differences of the participants, because each participant only tried one condition.

Experiment Design. Our experiment has one independent variable and two conditions, individual programming and pair programming. This setup is considered as a basic experiment design. The dependent variables can be quantitatively measured (no. of errors, no. of lines of code, code quality, programming time). The experiment can be replicable on other groups. To be able to repeat the same experiment on another domain easily. In this experiment we worked hard to remove any biases or any affect that may cause affect the result.

Task. The participants have to write a java program, to perform a binary calculation. The binary calculator will be used to add two positive 8 bits numbers. In this program, the users asked to input two positive 8 bits numbers. The program will validate the inputs, if it is correct, the program will produce the output.

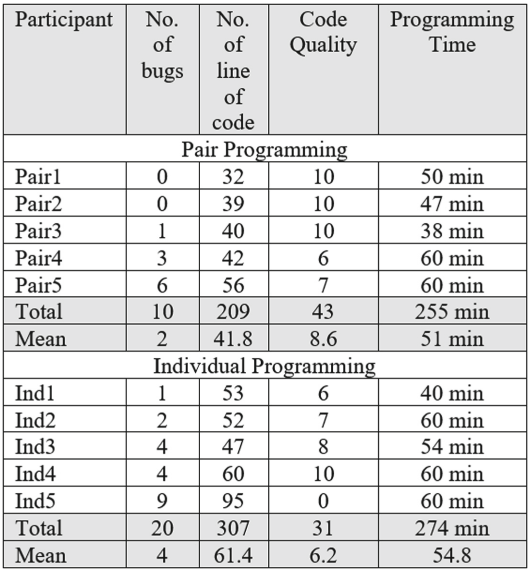
Experiment Procedure. At the beginning, we announced that we will hold a competition to develop a program in Java language. The top three winners will win a $100 prize for each one. We asked those who wish to participate, to register through the Google form, prepared specially for that. 15 participants, who registered first, were selected to do the task. The task was carried out through the following steps:

1. We conducted a pilot study, over one student as individual and two students as pair.
2. We evaluated the result taken from pilot study to make sure that the experiment's procedure we adopted work as required.
3. After 10 days, we called the 15 participants, to do the experiment.
4. We welcomed the participates and divided them into two groups, 5 individual groups and 5 pair groups based on their preferences.
5. We gave them a printed copy of the task with the details.
6. We introduced the task for participants, and we gave a brief information about pair-programming approach.
7. The participants who chose pairs were sat in the left row of the lab, and the individual groups were sat in the right row.
8. The participants were asked to submit the code into Google form to facilitate the process of transferring the code after completing the task.
9. The task was performed using Java language through eclipse IDE, in the university lab, to facilitate the process of compiling the code for participants.
10. We made the experiment easy and smooth and removed all biases.
11. The experiment was carried out one time; the participants were started at the same time, without knowing any description about the task.
12. The participants were asked to submit the code through Google form after they finished the task.
13. We thank all the students for their participation.
14. We sent the code to the SQA expert for analyzing and evaluating.
15. We took the data from SQA expert.
16. Finally, we evaluated and analyzed the data. Then we extracted the results

Data Collection. The data collected, by analyzing the programs where the participants submitted through Google form as shown in Table 1.

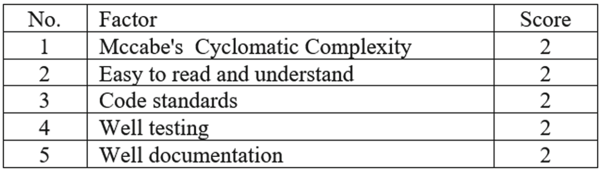
Data Analysis. The process of collecting the data taken from analyzed code generated by the participants after finishing the experiment. The number of errors for each participant has been calculated in addition to the number of lines

Table 1. The results taken by PP and Individual programming.



Participant	No. of bugs	No. of line of code	Code Quality	Programming Time
Pair Programming				
Pair1	0	32	10	50 min
Pair2	0	39	10	47 min
Pair3	1	40	10	38 min
Pair4	3	42	6	60 min
Pair5	6	56	7	60 min
Total	10	209	43	255 min
Mean	2	41.8	8.6	51 min
Individual Programming				
Ind1	1	53	6	40 min
Ind2	2	52	7	60 min
Ind3	4	47	8	54 min
Ind4	4	60	10	60 min
Ind5	9	95	0	60 min
Total	20	307	31	274 min
Mean	4	61.4	6.2	54.8

Table 2. Factors for evaluation the code quality



No.	Factor	Score
1	Mccabe's Cyclomatic Complexity	2
2	Easy to read and understand	2
3	Code standards	2
4	Well testing	2
5	Well documentation	2

of the generated code. But for the code quality, the code analyzed by an expert who put a score for each code based on the following factors [12]:

The SQA expert has split both codes written by two approaches (Individual, Pair) and judge based on the following factors:

1. Number of line of code taken from the IDE.
2. Number of bugs based on the expected bugs list.
3. High Quality score for each approach has taken from the total of factors mentioned in Table 2.
4. Programming time has been calculated based on the difference between start time and submitted time.

The expert, who was relied on to judge the code, finished a master degree in software engineering. He has 12 years of experience in application development using Java Language in addition to his expertise in software quality assurance. We used the independent samples t-test to do the analysis because the design of our experiment is between group, and there is one independent variable, and two conditions. We used IBM SPSS program to perform the statistical analysis.

4 Results

The main purpose of this study is to shed light on the most important results obtained from the executed experiment in terms of the following factors:

Number of Errors: To understand how pair-programming effect on reducing number of errors, we have prepared a list of 9 bugs that should the participant avoid it, as shown in Table 1. After the experiment executed and the code submitted through Google form, we tested the code and we observed the following results:

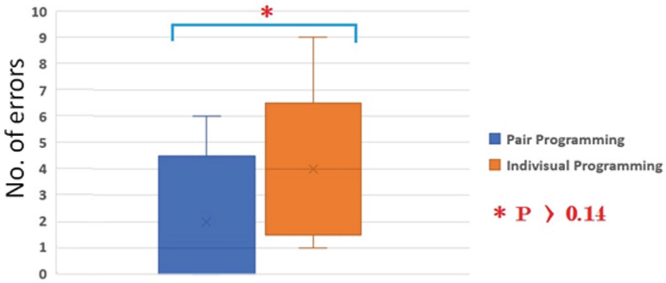


Fig. 2. Chart showed mean of number of errors for each group

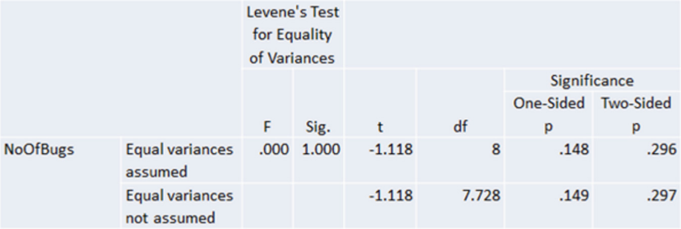
As shown in Fig. 2, we noticed that the mean for number of errors registered in pair programming less than the mean of errors registered from individual. In order to investigate the truth of null hypothesis (H0) that there is no significant difference between pair programming and individual programming in reducing number of errors, an independent samples t-test was conducted to prove this hypothesis. We clearly observed that the standard deviation (2.54951) in pair programming, while in individual the standard deviation (3.08221). The ratio of errors in pair-programming 50% fewer than individual.

The results in the Table 3 have shown that p-value $(f(0.00) = (1.000) > 0.05))$, and p-value of $(t(-1.118) = (0.148) > 0.05)$, that's mean, we can't reject the null hypothesis, because in science we want 95% to be significant, but here, there is a probability of 85.2% chance that the data random significant, which its refused.

Number of Lines of Code: In this suction, the effect of pair programming on the number of line of code clearly observed. The results below have shown the differences between both approaches pair programming and individual programming.

As shown in Fig. 3, we noticed that the mean for number of line of code registered in pair programming was less than the mean of number of line of code

Table 3. Independent sample t-test of number of errors



		Levene's Test for Equality of Variances					
						Significance	
						One-Sided	Two-Sided
		F	Sig.	t	df	p	p
NoOfBugs	Equal variances assumed	.000	1.000	-1.118	8	.148	.296
	Equal variances not assumed			-1.118	7.728	.149	.297

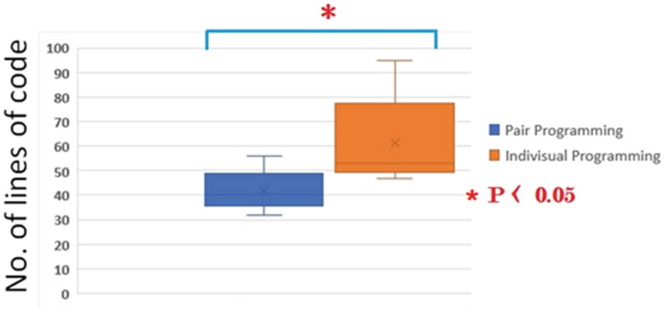


Fig. 3. Chart showed mean of number of lines of code for each group

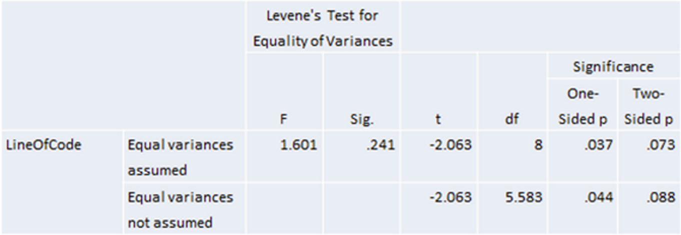
registered from the individual. In order to investigate the truth of null hypothesis (H0) that there is no significant difference between pair programming and individual programming in reducing number of line of code, an independent samples t-test was conducted to prove this hypothesis. The results clearly observed that the standard deviation was (8.78635) in pair programming, but in the individual, standard deviation was (19.34683). The ratio of number of line of code in pair-programming 30% fewer than individual.

The results in the Table 4 showed that p-value ($f(1.601) = (0.241) > 0.05)$), and p-value of ($t(-2.063) = (0.037) < 0.05$). This mean we reject the null hypothesis because the value significantly different. In science we want 95% or more to be significant, but here, there is a probability of 96.3% chance that the data random significant. So, the hypothesis was rejected.

Code Quality: As known that code quality is not quantitative factor, but it is qualitative, and it is difficult to measure, therefore, we have adopted a mechanism, through which we can give numerical values for the factors related to the quality of the code as mentioned in Table 2.

As shown in Fig. 4, we noticed that the mean for quality code registered in pair programming more than the mean of the quality code registered from the individual. In order to investigate the truth of null hypothesis (H0) that there is no significant difference between pair-programming and individual programming

Table 4. Independent sample t-test of number of line of code



		Levene's Test for Equality of Variances					
						Significance	
		F	Sig.	t	df	One-Sided p	Two-Sided p
LineOfCode	Equal variances assumed	1.601	.241	-2.063	8	.037	.073
	Equal variances not assumed			-2.063	5.583	.044	.088

in building high quality code. An independent samples t-test was conducted to prove this hypothesis.

We clearly observed that the std. deviation was (1.94936) in pair programming, but in the individual, std. deviation was (3.76829). The ratio of building high quality code in pair-programming 25% more than individual.

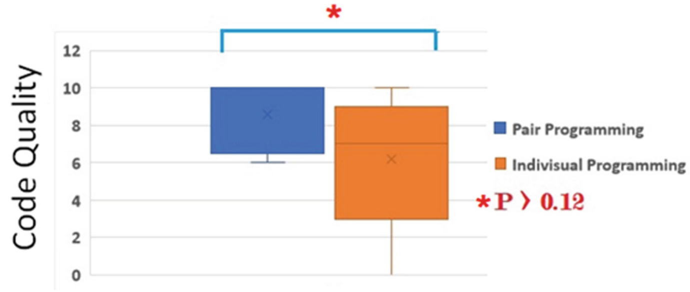


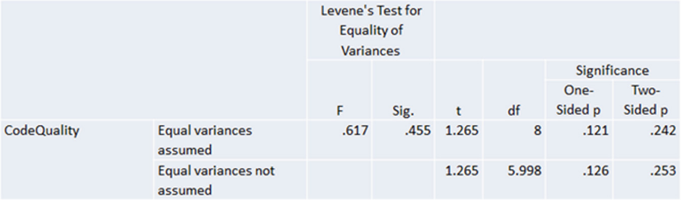
Fig. 4. Chart showed mean of quality code for each group

The results in the Table 5 showed that p-value (f(0.617) = (0.455) > 0.05)), and p-value of (t(1.265) = (0.121) > 0.05), that's mean, we absolutely not reject the null hypothesis, because, the value of both group is not significantly different. In science we want 95% or more to be significant, but here, there is a probability of 87.9% chance that the data random significant.

Programming time: We have noted that the time spent in the implemented experiment in both approaches is very close. It may be due to the fact that the program is small, and it is difficult to show significant differences in the programming time. Perhaps also due to the loss of time due to the discussion among team members in the pair programming approach.

As shown in Fig. 5, we noticed that the mean for programming time registered in pair programming less than the mean of time registered from individual. In order to investigate the truth of null hypothesis (H0) that there is no significant

Table 5. Independent sample t-test of quality code



		Levene's Test for Equality of Variances					
						Significance	
		F	Sig.	t	df	One-Sided p	Two-Sided p
CodeQuality	Equal variances assumed	.617	.455	1.265	8	.121	.242
	Equal variances not assumed			1.265	5.998	.126	.253

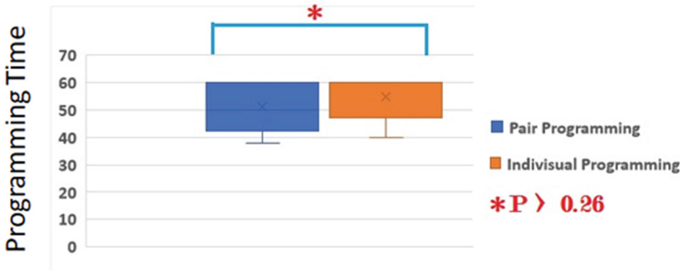
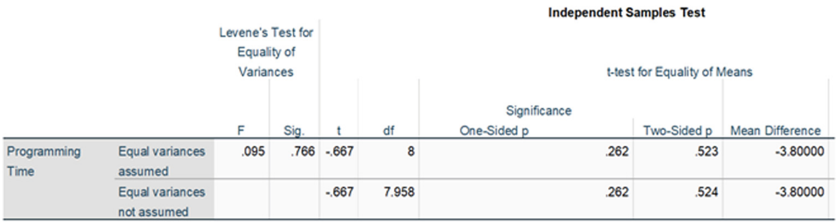


Fig. 5. Chart showed mean of programming time for each group

difference between pair programming and individual programming in reducing programming time, an independent samples t-test was conducted to prove this hypothesis.

Table 6. Independent sample t-test of number of line of code



Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
						Significance		
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference
Programming Time	Equal variances assumed	.095	.766	-.667	8	.262	.523	-3.80000
	Equal variances not assumed			-.667	7.958	.262	.524	-3.80000

The results in Table 6 revealed that the p-value ($f(0.095) = (0.766) > 0.05$)), and p-value of ($t(-0.667) = (0.262) > 0.05$), that's mean, we can't reject the null hypothesis, because in science we want 95% to be significant, but here, there is a probability of 73.8% chance that the data random significant, which its refused.

5 Discussion

As shown in the results, the programs that produced by pair programming has significant differences compared with the individual programming. In this section we will explain why we got the following results: **Less number of errors:**

- Once the driver has finished writing the statement of code, the navigator reviewed and tested the code being written. So, lot of errors was caught as they are being typed.
- The expertise of both members enabled them to do their best work, which helped eliminating errors. The idiom says "Two heads better than one head".
- The code being looked by both pair programming members, which helped to reduce the errors.
- Using different use cases by both members helped in reducing number of errors, because they applied full coverage paths testing.

Less Number of Lines of Code:

- The driver and navigator were able to jointly reduce some unnecessary statements.
- The navigator was able to suggest ways to reduce the number of code statements without affecting on program functionality.
- Both members were able jointly to use different design pattern, which helped to reduce number of lines of code.

High Quality Code:

- The expertise of both members can help avoiding code complexity by reducing number of paths.
- The navigator looks at the code from a different point of view. He can direct the driver to write an understandable code in case the code is ambiguous.
- Navigator can guide the driver to use code standards to avoid different definitions and terminologies.
- When the driver writes the code in heterogeneous context, the navigator can direct him to follow a consistent style.

Minimizing Programming time: There were no significant differences in the programming time between both approaches. That's due to lost time during discussion between members, and also because the program is small, which made the programming time close between them.

Observations:

- The effect of pair programming on maintainability will be more evident when pairs are systematically formed.
- There is a big difference in the results of some pairs, perhaps the reason for this was the differences of their expertise.
- In order to obtain more accurate results, the levels of the participants must be evaluated before forming pairs and individuals.
- The Implementation of the experiment in a competitive manner among the participants was the reason for the success of the experiment.

6 Conclusion and Future Work

In this study, we focused on the effect of pair programming on code maintainability, in terms of number of errors, number of lines of code and the code quality. Most reviewed studies talking about pair programming on educational domain in terms of collaborative, sharing knowledge and confidence between students. Some universities used it as an effective coding education method, to improve the quality of education. The conducted experiment in this research applied on educational domain by a small sample of novice students. The results clearly revealed that there is a significant difference in building program with 50% less errors, 30% less code, and 25% better code. Thus, there is an effect on code maintainability. Furthermore, there is a marginal reducing in programming time. In the future, we will conduct an experiment in different domains with large sample and different experiences of participants, to show if there more effect on code maintainability.

References

1. Misra, S.: Pair programming: an empirical investigation in an agile software development environment. In: Przybyłek, A., Miler, J., Poth, A., Riel, A. (eds.) LASD 2021. LNBIP, vol. 408, pp. 195–199. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-67084-9_13
2. Sunitha, K.S., Nirmala, K.: Correlation study on defect density with domain expert pair speed for effective pair programming. Indian J. Sci. Technol. **8**(34), 1–7 (2015)
3. Rajagopal, S., Hareesha, K.S., Kundapur, P.P.: Optimising pair programming in a scholastic framework: a design pattern perspective. J. Comput. Sci. **13**(6), 199–210 (2017)
4. Nawahdah, M., Taji, D., Inoue, T.: Collaboration leads to success: a study of the effects of using pair-programming teaching technique on student performance in a Middle Eastern society. In: 2015 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE). IEEE (2015)
5. Anjum, S., Batul, H., Sirshar, M.: A comparative study of ensuring quality in pair programming. IJCCSE **2**(1), 160–171 (2015)
6. Alazzam, I., Akour, M.: Improving software testing course experience with pair testing pattern. Int. J. Teach. Case Stud. **6**(3), 244–250 (2015)
7. Celepkolu, M., Boyer, K.E. The importance of producing shared code through pair programming. In: Proceedings of the 49th ACM Technical Symposium on Computer Science Education (2018)
8. Ersoy, I.B., Mahdy, A.M.: Agile knowledge sharing. Int. J. Softw. Eng. (IJSE) **6**(1), 1–15 (2015)
9. Choudhari, J., Suman, U.: An empirical evaluation of iterative maintenance life cycle using XP. ACM SIGSOFT Softw. Eng. Notes **40**(2), 1–14 (2015)
10. Sadath, L., Karim, K., Gill, S.: Extreme programming implementation in academia for software engineering sustainability. In: 2018 Advances in Science and Engineering Technology International Conferences (ASET). IEEE (2018)
11. Tsompanoudi, D., et al.: An empirical study on factors related to distributed pair programming, 65–81 (2019)

12. Bogner, J., Wagner, S., Zimmermann, A.: Automatically measuring the maintainability of service-and microservice-based systems: a literature review. In: Proceedings of the 27th International Workshop on Software Measurement and 12th International Conference on Software Process and Product Measurement (2017)
13. Chen, K., Rea, A.: Do pair programming approaches transcend coding? Measuring agile attitudes in diverse information systems courses. J. Inf. Sys. Educ. **29**(2), 53–64 (2018)

Relevant Knowledge Use During Collaborative Explanation Activities: Investigation by Laboratory Experiment and Computer Simulation Using ACT-R

Yugo Hayashi[1](✉) and Shigen Shimojo[2]

[1] College of Comprehensive Psychology, Ritsumeikan University, 2-150 Iwakura-cho,
Osaka, Ibaraki 567-8570, Japan
y-hayashi@acm.org

[2] Graduate School of Human Science, Ritsumeikan University, 2-150 Iwakura-cho,
Osaka, Ibaraki 567-8570, Japan
cp0013kr@ed.ritsumei.ac.jp
http://www.hayashilab.jp/en/

Abstract. This study investigated the nature of knowledge use during collaborative explanation activities through a laboratory experiment and a computer simulation. We focused on how relevant knowledge is used in individual learning and how the use of such individual knowledge, influenced by cognitive bias (e.g., misconceptions), affects explanation activities and group activities in developing a concept map. Our results show that learners used relevant knowledge retrieved from memory during group activities, but some learners changed their strategy to use different knowledge during explanation activities. We further examined the cognitive process by developing an Adaptive Control of Thought-Rational (ACT-R) model. We focused on the knowledge retrieval process of self/other's knowledge brought about by discrepancies in perspectives. The simulation results showed that the opportunity to search for knowledge brought a chance to retrieve relevant knowledge for generating adequate explanations in collaboration exercises.

Keywords: Collaborative learning · Explanation · Misconceptions · Concept map · Computer simulation

1 Introduction

Learning by explanation has been studied in the fields of cognitive science and learning science. It has been reported as an efficient strategy for gaining an understanding of the learning material and is effective in triggering metacognition [2,4]. Deeper understanding may be possible by generating abstract concepts and generalizations [12,26]. Studies also showed that explanations to oneself facilitate generalizations to solve problems [20]. Studies in educational settings,

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 52–66, 2022.
https://doi.org/10.1007/978-3-031-20218-6_4

such as active learning, adopt explanation activities for collaborative learning activities [23]. It has been shown that explanations of others may facilitate constructive interactions for the development of knowledge [21,24]. As mentioned in the Integrated Cognitive Antisocial Potential (ICAP) theory [3], constructive and interactive activities in collaboration with peer learners are triggered by explanation activities wherein learners actively argue with other claims and perspectives. However, social group studies and collaborative problem-solving studies show that individuals may make explanations based on misunderstandings, and discussions may be misdirected due to naïve knowledge and biased perspectives [13,27]. Developing common ground and mutual understanding [5,11]is key for linguistic interpretations during collaborative activities; however, the cognitive mechanisms that underlie this process are still unclear. Thus, it is necessary to understand the process of relevant and irrelevant knowledge use during explanations in collaborative activities to develop cognitive models further. This study conducted an experiment observing how dyads use adequate, relevant knowledge during explanation activities using concept maps. We also developed an Adaptive Control of Thought-Rational (ACT-R) cognitive model to explain how individuals adopt knowledge during explanation activities.

1.1 Knowledge Use During Collaborative Learning: Misconceptions and Conformity

Past studies on collaborative problem-solving show that explaining is an effective strategy to gain a better understanding of the topic at hand. However, some negative aspects hinder adequate reasoning and knowledge use. Individuals prefer to use prior knowledge by retrieving information from knowledge learned in the past [8] , and sometimes, they may include naïve and false beliefs [13]. Moreover, individuals have a cognitive tendency to use accessible knowledge, which can include biases [18]. Explanation activities have been reported to facilitate a better understanding of misbeliefs because making mechanistic explanations may constrain casual inferences by reducing biased or misbeliefs [20]. However, novice learners who have a poor ability to self-monitor their knowledge use during explanation activities may use misconceptions to generate inferences. Another difficulty in collaborative settings is the understanding of how relevant and irrelevant knowledge is shared among groups. During explanation activities, an individual will perceive the knowledge of others and will interpret his own knowledge to make sense of the explained content [19]. From a cognitive processing perspective, searching for knowledge through memory retrieval may not always be successful and may retrieve inadequate knowledge, leading to discrepancies in knowledge sharing [15,16]. Moreover, misconceptions may lead to confusion and irrelevant memory and knowledge retrieval [6]. When learners confront different perspectives, they must decide which knowledge should be used, and this may involve social aspects such as decision-making or social influences. Many studies in social psychology have reported the negative effects of social influences, such as conformity, hindering the process of adequate decision-making during problem-solving

[7]. Relevant knowledge use during explanation is difficult and influences collaborative learning activities for students. However, only a few studies have investigated how group members use relevant knowledge at the individual and group level. Many past studies have focused on how the use of appropriate knowledge and the integration of different perspectives lead to success. However, little is known about the use of relevant or irrelevant knowledge, how they are retrieved, and how they are used in interactions with other members. Moreover, cognitive models of the individual use of memory/knowledge and how they are used in collaborative learning have not been developed.

1.2 Cognitive Model in Collaborative Learning

Several computational models of collaboration and coordination have been developed in cognitive science. [10] investigated how communication systems emerge during the communication process. Modeling using machine learning and sensing technology has also been recently attempted. For example, synchrony between individuals during collaborative work has been used to evaluate the success of communication [25]. Recurrence analysis is another line of study focusing on modeling the communication process in collaborative activities [22]. However, little attention has been given to modeling the process of how dyads use their respective knowledge to develop a shared understanding in collaborative problem-solving. [14] used cognitive task analysis and production rules to see how individuals with different perspectives and knowledge share their common knowledge. However, this study did not consider how the individual's prior knowledge and its influence on explanation activities. Moreover, further investigation must be conducted considering the effects of individual biases, the success of knowledge retrieval and associate learning, and the misbeliefs during problem-solving, all of which bring discrepancies and confusion in collaborative activities. Success in making adequate inferences during explanation activities towards others can be influenced by errors during memory retrieval and the type of knowledge retrieved. [16] showed that inadequate retrieval from memory of a certain problem and its corresponding instances would cause misunderstanding about knowledge use. The naive use of memory has also been studied as an influence to the inferences of novice learners [13]. Additionally, collective inferences and social dynamics, including instances of conformity, may influence the decision to use retrieved knowledge [28]. Thus, the following points should be investigated in developing cognitive models: (1) memory retrieval or non-retrieval (perspective taking), (2) failure in memory retrieval, (3) use of naïve memory, and (4) conformity by adopting a partner's perspective. Thus, this study focuses on modeling an individual's use of relevant memories during collaborative learning and how it influences developing mutual knowledge. The present study uses the ACT-R architecture [1], which can assess the control and flow of the information process of relevant knowledge use. Relevant and irrelevant knowledge can be harvested in the form of declarative memory in the ACT-R. Productions can be implemented for retrieval and search of relevant parameters that elaborate on cognitive biases and noises that may cause failure to use such knowledge.

1.3 Goal and Hypothesis

This study focuses on a collaborative learning process in which learners constructively interact with each other by making explanations using a concept map (CM). We investigated how relevant knowledge is used from individual learning and how the experience of using such knowledge influences the CM development in a group consisting of members with different knowledge. Thus, the first goal of this study is (1) to experimentally investigate how individuals use relevant knowledge retrieved from memory during explanation activities and further explore how it influences collaborative performance. The second goal is (2) to develop a cognitive model of knowledge retrieval during collaboration and conduct a computer simulation based on protocols from experimental results. We propose the following three hypotheses for the experiments conducted.

H1: Explanation activities facilitate the use of inference strategies, such as the use of relevant knowledge retrieved from memory.
H2: Prior experience of an individual's knowledge retrieval strategy may influence conversations and inference strategies on group outcomes.
H3: Change in the knowledge retrieval strategy may occur because of the discrepancies caused by group interactions.

2 Experimental Process

2.1 Participants

This study used data from a previous anonymized study of 60 university students majoring in psychology that participated in a laboratory-based experiment for course credit. Hereafter, we refer to these participants as the learners.

2.2 Experimental Task Procedure

This study focuses on dyads involved in an explanation-based activity [29] conducted in a remote environment that allowed them to communicate with each other using a concept map tool [9]. After entering the laboratory, the dyad of learners was introduced to each other and was given instructions about informed consent. This study was ethically reviewed by the relevant institutional review board. The learners' goal in this experiment was to explain a particular case using psychological theories. The experiment included the following three phases: (1) individual text learning phase, (2) individual concept map-generation phase, and (3) collaborative reasoning phase, wherein they gave a shared representation of their concept maps. The individual learning phase consisted of two sub-phases: (a) memorizing the theoretical concept of attribution theory and (b) memorizing the case story. In the individual learning phase, they were required to apply the attribution theory from [29] to a case of a student who participated in a school counseling program, and they describe why the student had anxiety about the new academic year. In the individual concept map generation phase, they were required to create a concept map to explain the case using theoretical knowledge. In the concept map, they were allowed to use different types of links to connect the nodes to describe the attribution process. Nodes used in the task were nouns in the case story, e.g., “student,” “counselor,” “school,” “math grade,” “semester.” To connect the nodes with the links in the concept map, the learners would use knowledge from the attribution theory such as “internal cause” and “external cause.” An adequate strategy to complete the concept map was to use the words from the learning text as links and connect nouns from the case theory. For example, when a learner tries to explain the situation in the case story, such as the student complained to the counselor that he obtained a bad grade in math last semester because his parents were not good at math. This can be expressed by using two nodes, node1 (“math grade”), node2 (“parents”), and a link (“internal cause”). Using the relevant knowledge about the links is the key to theoretically explaining the situation in this task. In the collaborative reasoning phase, learners worked in pairs by discussing the same task they worked on in the individual concept map phase. As shown in Fig. 1, two monitors were connected to the PCs, and the Cmap software (https://cmap.ihmc.us/) was installed on the PCs to develop and synchronize concept maps. This set-up allowed for the simultaneous production and sharing of concept maps, thereby enabling each learner to see the other’s concept. This also allows them to develop a new concept map together.

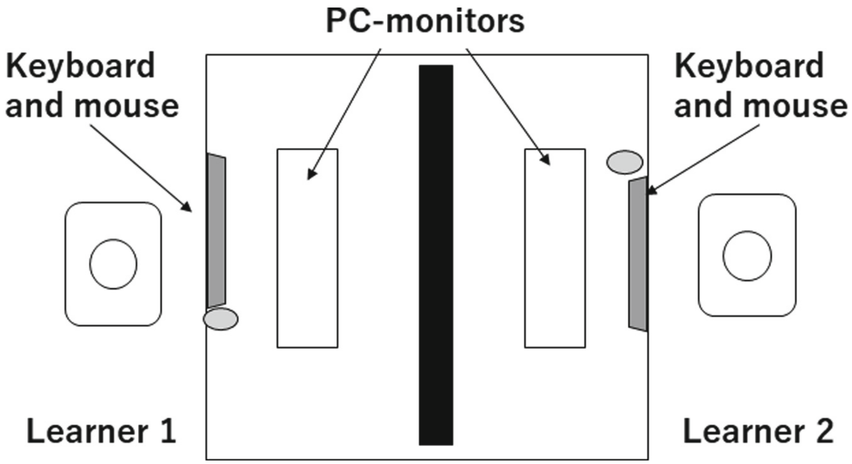


Fig. 1. Experimental set-up.

In the collaborative learning phase, the learners worked together by providing oral explanations. They were instructed to explain each other’s thoughts to develop a new concept map. As described in Fig. 2, the participants were able to see each other’s concept map (right-hand side) developed in the individual learning phase while working on the shared concept map (left-hand side). The

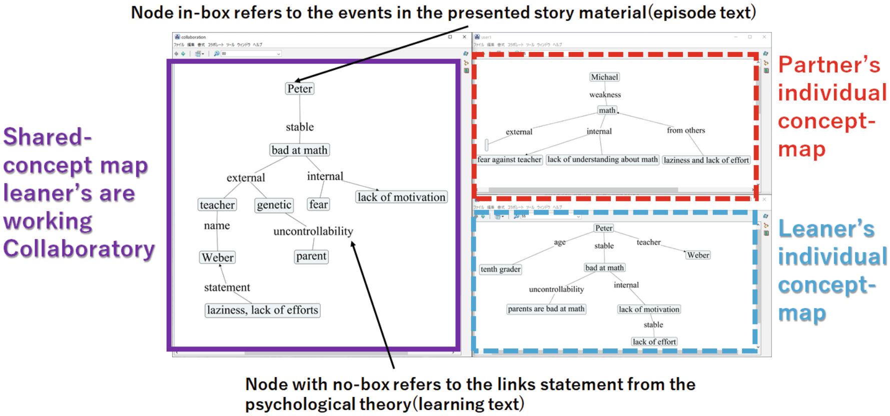

Node in-box refers to the events in the presented story material(episode text)
Shared-concept map leaner's are working Collaboratory
Peter
stable
bad at math
external
internal
lack of motivation
teacher
genetic
fear
name
uncontrollability
Weber
parent
statement
laziness, lack of efforts
Partner's individual concept-map
Michael
weakness
math
external
internal
from others
fear against teacher
lack of understanding about math
laziness and lack of effort
Leaner's individual concept-map
Peter
age
stable
teacher
Weber
tenth grader
bad at math
uncontrollability
internal
parents are bad at math
lack of motivation
stable
lack of effort
Node with no-box refers to the links statement from the psychological theory(learning text)

Fig. 2. Example screen of a learner in the collaborative learning phase. The right-hand side shows two windows of the previously generated concept maps at the individual phase. The left-hand side shows the shared concept map.

learners discussed and created a new concept map using mouse keyboards. This phase was allotted 15 min.

2.3 Measures

Relevant Use of Knowledge (Words) During Task Work: Adequate Links. A relevant strategy in the development of a concept map in this task is to use the knowledge acquired in the learning phase, which is of the learning text ("attribution theory"). However, as discussed previously, individuals have a cognitive bias in using individual or naïve knowledge to make inferences. To formalize this point, we analyzed all the text data generated in the concept map (all nodes and links) for each learner in each phase. Then, we coded each learner's text and calculated the ratio of knowledge (text) based on memory retrieval from the learning text and episode. The relevant link use rate was calculated by multiplying the number of links used from the learning text to the total number of links used by one learner. For each learner, we calculated this as the relevant retrieval rate.

Relevant Use of Knowledge (Words) During Collaboration: Verbal Protocol. Verbal data was collected during the collaborative reasoning phase for the analysis of the interaction process. All verbal data were textualized and processed by morphological analysis to code noun phrases. We coded nouns that were used based on memory retrieval from the learning and episode texts. Values were obtained as follows:

Relevant phrase use rate for learning text = phrase used from learning text /all phrases used

3 Results of Human Experiment

3.1 Use of Relevant Knowledge by Memory Retrieval at the Individual and Collaborative Phases

Measures for analyzing the relevant use of knowledge during a learner's activity in developing concept maps were used to analyze the data. Figure 3 shows the memory retrieval rate when using relevant words from the learning text (links).

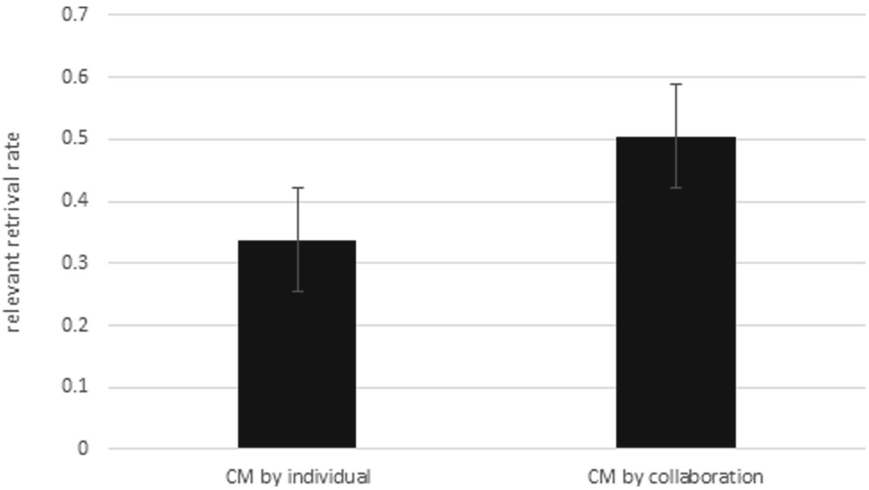


Fig. 3. Ratio of the use of relevant knowledge in the individual phase and in the collaborative phase.

First, the overall relevant retrieval rate shows that learners do not use the relevant words learned from the learning text. The average retrieval rate was less than 51%, which indicates that they used naïve knowledge. This supports our hypothesis of the existence of cognitive bias in using naive knowledge, as discussed previously. A one-way ANOVA was performed by phase (individual vs. collaboration) as mixed data factor analysis was conducted to further investigate how memory retrieval changed in each phase. There was a significant difference between the two factors $F(1, 59) = 8.222, p = 0.001, \eta^2 = 0.1223$). The results show that the relevant retrieval rate increased when learners collaborated compared to individual learning. This indicates that explanation activities during collaboration facilitated learners to use memory retrieval strategies, especially from the text. This result supports Hypothesis H1.

3.2 Influence of Prior Experience and Interaction on Inference Strategy in Collaborative Group Outcomes

Results from previous analyses show that prior experience of using relevant memory retrieval in the individual phase may have influenced the collaborative outcomes of the concept map. Moreover, strategies that emerged during the group interactions may also influence the outcome. Considering these points, we investigated the correlation of the relevant use of words used from the learning text for each of the three phases (individual, collaboration, and group dialogue). Table 1 shows the results of the correlation using Pearson's coefficient index.

Table 1. Correlation of relevant use of learning text (links) and phases (individual, collaboration, and dialogues).

	CM by individual	CM by collaboration	Dialogue by collaboration
CM by individual	–	.328*	0.209
CM by collaboration		–	.518**
Dialogue by collaboration			-

$*$ indicates significance by $p < 0.01$
$**$ indicates significance by $p < 0.001$

Next, we conducted an analysis of the learning text. For CM individual to collaborative (dependent variable), the regression coefficient $R^2$was 0.092, and the F-value from ANOVA was 6.989, indicating significance for both variables ($p = .011$). For dialogue to CM by collaboration (defendant variable), the regression coefficient R^2 was .269, and the F-value from an analysis of variance (ANOVA) was 21.306, indicating significance for both variables ($p = 0.000$). These results suggest that (1) prior experience in using knowledge influences group output and (2) type of knowledge use during interaction will influence the output, but (3) prior experience will not influence knowledge use during an interaction. This indicates that H2 is partially supported. Moreover, the evidence that learners did not use relevant knowledge in the dialogue used in the individual phase suggests that learners may shift to a different knowledge retrieval strategy. This supports H3, which will be further investigated in the following simulation.

3.3 Discussion

These results show that prior experience of individuals' inference strategy may influence group outcomes but not collaboration dialogues. Moreover, group explanation activities (captured by the dialogue) facilitate the use of relevant memory for the CM through collaboration. This suggests that some learners may decide not to use the relevant memory retrieval strategy during the interaction. Focusing on group members who used different inference strategies in their individual activities, they would see the differences in each other's strategies in the group collaboration phase. Past studies on collaborative problem solving [16] show that when members encounter discrepancies in their perspectives, members search for others' suggested perspectives. When they confirm that such perspective conformity occurs, they keep using their original and different perspectives. In a situation where learners shared different perspectives (nodes and links), they might have searched for the knowledge in their knowledge base to confirm their partner's varying perspective. To further investigate the learners who have confronted members with different knowledge or strategies and subsequently changed their strategy, we categorized learners into six groups depending on the knowledge used in each phase. Table 2 shows the results of the number of learners for each learning text (link). Numbers were assigned to each dyad for the individual phase (2 = both relevant, 1 = part relevant, 0 = both irrelevant) and the collaboration phase (1 = relevant, 0 = irrelevant). As for (2–1), both used relevant knowledge at the individual and group phases in developing the CM. In (1-1), one of their partners used relevant knowledge from the individual and eventually used the relevant knowledge in the group. In (0–1), one partner used irrelevant knowledge (have different perspectives) at the individual phase but decided to switch strategies to use relevant knowledge in the group phase. In (2–0), both will use relevant knowledge at the individual but irrelevant knowledge in the group phase. In (1–0), one of their partners used relevant knowledge at the individual phase, but both group members eventually failed to use relevant knowledge in the group. In (0-0), both used irrelevant knowledge at the individual phase, and both failed to use relevant knowledge in the group. The average score (shown in Fig. 3) was used as the criterion for learners using relevant or irrelevant knowledge. Learners in (1-1) and (1–0) are the groups that were defined as members who have partners who used different knowledge in the individual phase. We further examine these pairs in the next section.

Table 2. Number of success and failure groups to generate relevant link (learning text) during each session (single/pair).

		CM by collaboration	
		1(relevant)	0(irrelevant)
CM by individual	2(both relevant)	6	2
	1(part relevant)	6	5
	0(irrelevant)	5	6

A chi-square analysis was also conducted, but there was no significant difference (χ^2 (2) = 1.678, *n.s.*); however, interestingly, several pairs (11/30) used different memory retrieval strategies. The results in Tables 1 and 2 show that learners who used both use the same relevant or irrelevant knowledge that an individual tends to use the same knowledge in the group, which is consistent with the results of the correlation analysis. Notably, several learners were using different knowledge and strategies, and one of them changed their strategy of knowledge use. Learners might have changed their strategy by simply using their partners' knowledge or searching for the partner's different knowledge observed from the CM. Moreover, the opportunity to search for knowledge might have brought about the chance of successfully retrieving relevant knowledge. This cognitive process is difficult to capture from our data; therefore, we developed a computer simulation model using ACT-R and explored whether the opportunity to search for knowledge may influence knowledge retrieval.

4 Computer Simulation by ACT-R

We used the ACT-R architecture to model the knowledge use explained in the previous section [1]. To further investigate the point suggested in our previous section, we investigated the process of knowledge use when learning to confront a partner's different knowledge. We developed a model of the learner who has confronted discrepancy and succeeded, such as is the case for (1-1) mentioned in the previous section. The basic flow of the model articulated in ACT-R was as follows: (1) look at the partner's CM, (2) search for their knowledge, (3) identify if they use the same or different knowledge, (4) search for knowledge in the learned text of self or other, and (5) input knowledge for group CM. Figure 4 shows the overall flow of the model implemented in ACT-R.

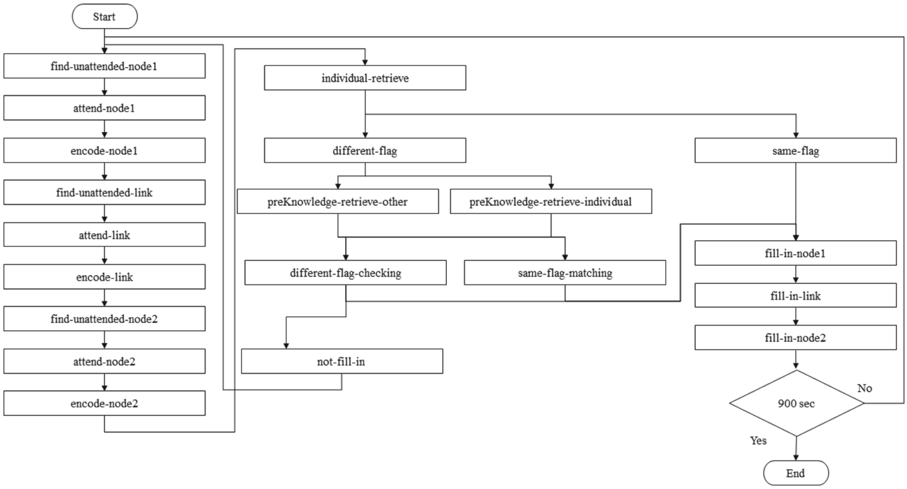


Fig. 4. Overall flow of the model in the (1-1) strategy implemented in the ACT-R.

In the (1-1) model, an individual finds a difference and retrieves knowledge from the learning text to confirm if it matches their current posed knowledge. The individual then uses the most activated knowledge. The parameters used for knowledge retrieval and data fitting are described in the next section.

4.1 Parameters

The goal of this simulation is to determine the mechanism of (a) decision and selection about self/other knowledge and (b) how success or failure in the retrieval of relevant memory influences the use of knowledge for CM in a group setting. For the process in (a), which was the selection of which self/other knowledge will be used for the search (preKnowledge-retrieve-other/individual), we used the utility parameters in ACT-R. The probability of selection was calculated using the following equation:

$$Probability(i) = \frac{e^{i/\sqrt{2s}}}{\Sigma_i e^{U_i/\sqrt{2s}}} \tag{1}$$

where U_i is the index of expected utility. The reference count N in the utility is expressed as $U_j(n)$.

$$U_i - (n) = U_i - (n-1) + \alpha[R_i(n) - U_i(n-1)] \tag{2}$$

where i is the production. For the second part of the modeling process, which was the memory retrieval of learning text (chunk) in the process in (b), we used the activation rate of memory retrieval. The equation used for activation is as follows:

$$A_i = B_i + \varepsilon \tag{3}$$

where B_i is the base level of activation, ε is the noise (same as utility).

$$B_i = ln\frac{n}{1-d} - d * ln(L) \tag{4}$$

$$\sigma^2 = \frac{\pi^2}{3}s^2 \tag{5}$$

where n is the number of presentations of chunk i, L is the lifetime of chunk i (the time since its creation), and d is the decay parameter (the value of bll) = 0.5.

4.2 Fitting Data

The fitting data (dependent variable) are the relevant knowledge retrieval rates of the learning text. The average retrieval rate for the link was 0.368 (individual) and 0.697 (collaboration) in (1-1).

4.3 Fitting to Experimental Data

By fitting the data using the model, we set the parameters that best fit the experimental data. The default values were used as the parameter for the activation of the chunk, that is, learning text. For the utility parameters, we looked for the best fit with the activation rate by adjusting the parameters to 8.0, 9.0, 10.0, and 12.0. Through these attempts, we found that the utility parameter that best fit the activation rate for the links was 9.0. The activation rate using these parameters in the collaboration phase was 0.611.

4.4 Investigating the Influence of Searching for Self/Other Knowledge Using the Model

We next conducted a simulation by comparing this model with a model that does not have the function of searching for partners' knowledge to replicate the individual phase. More specifically, this model was articulated by removing the production rule of (a) explained in the parameters section. The results are consistent with human data, as shown in Fig. 5, which demonstrates the findings of simulation data using the model. More specifically, the relevant retrieval rate increased from the individual phase to the collaboration phase, similar to the human experimental data. This can be due to the difference in the function (production rules) to search for the partner's perceived knowledge.

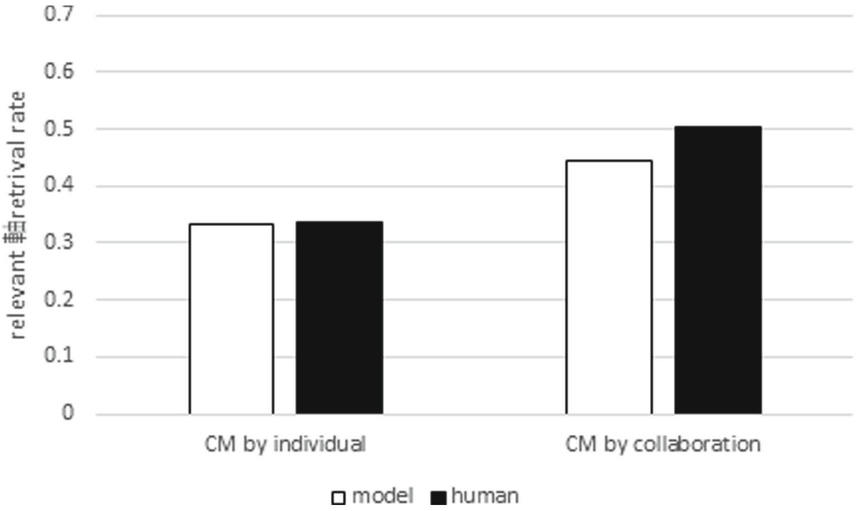


Fig. 5. Comparison between the relevant retrieval rate of the model and the human data implemented in the ACT-R.

4.5 Discussion

The computer simulation results using ACT-R show that our model was successful in replicating the results of the experiment. This model focuses on learners confronted with members having different perspectives and how learners will search for adequate knowledge to address the discrepancy. This opportunity to search could have functioned as crosschecking of one's knowledge and led to the realization of misconceptions. Searching for knowledge can be interpreted as monitoring and may be part of the metacognitive process, which has been shown to play an important role during collaborative learning activities [2]. Importantly, the decision to search for others' knowledge was brought about by the discrepancy of different perspectives [16]. Therefore, using different knowledge from individual experiences may play an essential role in the emergence of knowledge retrieval. This must be further investigated by developing a broader model that includes such a process. However, this can depend on the frequency of how learners bring up each other's different knowledge during their conversations. Moreover, behavioral evidence such as mouse clicks and gaze patterns can be further investigated to see how biased attention to knowledge influences decision-making. This can be further investigated by collecting data using eye-tracking devices such as in [17].

5 Conclusion

This study investigated collaborative learning in which learners constructively interacted by making explanations using a concept map (CM). We investigated how relevant knowledge is used from individual learning and how the experience of using such knowledge influences explanation activities in developing collaborative CMs with partners. This study first experimentally investigated the nature of how individuals use relevant knowledge retrieved from memory during explanation activities and explored how it influences collaborative performance. The experiment results show that explanation activities facilitate inference strategies, such as the use of relevant knowledge retrieved from memory. Moreover, prior experience of an individual's knowledge retrieval strategy influenced the relevant knowledge use during explanation activities and interactions at the group level. Some learners changed their strategy of knowledge retrieval during an interaction. To investigate this, we conducted a computer simulation using ACT-R to explain the cognitive process. The cognitive model focused on searching for the relevant knowledge of self/other, brought about by the discrepancy in self/others knowledge in the CM experiment. The computer simulation results show that our model and its parameters captured the relevant knowledge retrieval rate in such situations. Our model also explained how relevant or irrelevant knowledge might be used. The present study has important implications on the cognitive processes of how learners use misconceptions during explanation activities and how learners may change perspectives during explanation activities in collaborative learning.

Acknowledgments. This work was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant-in-Aid for Scientific Research (B), Grant Number JP20H04299.

References

1. Anderson, J.R.: Human symbol manipulation within an integrated cognitive architecture. Cogn. Sci. **29**(3), 313–341 (2005). https://doi.org/10.1207/s15516709cog0000_22

2. Chi, M.T.H.: Self-explaining expository texts : the dual processes of generating inferences and repairing mental models. In: Advances in Instructional Psychology : Educational Design and Cognitive Science, vol. 5, pp. 161–238 (2000). https://cir.nii.ac.jp/crid/1571135651239648768
3. Chi, M.T.H.: Active-constructive-interactive: a conceptual framework for differentiating learning activities. Top. Cogn. Sci. **1**(1), 73–105 (2009). https://doi.org/10.1111/j.1756-8765.2008.01005.x, https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1756-8765.2008.01005.x
4. Chi, M.T.H., Deleeuw, N., Chiu, M.H., Lavancher, C.: Eliciting self-explanations improves understanding. Cogn. Sci. **18**(3), 439–477 (1994). https://doi.org/10.1016/0364-0213(94)90016-7
5. Clark, H.H., Brennan, S.E.: Grounding in communication. In: Resnick, L.B., Levine, J.M., Teasley, S.D. (eds.) Perspectives on socially shared cognition, pp. 127–149. American psychological association (1991). https://doi.org/10.1037/10096-006. conferece on socially shared cognition, University of Pittsburgh, Learning Res & Dev Ctr, Pittsburgh, PA, FEB, 1989
6. Clement, J.: Overcoming students' misconceptions in physics fundamental change in children's physics knowledge. J. Res. Sci. Teach. **28**, 785–797 (1987)
7. Deutsch, M., Gerard, H.B.: A study of normative and informational social influences upon individual judgment. Comput. Educ. **51**(3), 629–636 (1955). https://doi.org/10.1037/h0046408
8. Dunbar, K.: How scientists really reason: scientific reasoning in real-world laboratories, pp. 365–395. MIT Press, Cambridge (1995). https://doi.org/10.1037/h0046408
9. Engelmann, T., Hesse, F.W.: Fostering sharing of unshared knowledge by having access to the collaborators' meta-knowledge structures. Comput. Hum. Behav. **27**(6), 2078–2087 (2011). https://doi.org/10.1016/j.chb.2011.06.002
10. Galantucci, B., Sebanz, N.: Joint action: current perspectives. Top. Cogn. Sci. **1**(2), 255–259 (2009). https://doi.org/10.1111/j.1756-8765.2009.01017.x
11. Garrod, S., Anderson, A.: Saying what you mean in dialog - a study in conceptual and semantic coordination. Cognition **27**(2), 181–218 (1987). https://doi.org/10.1016/0010-0277(87)90018-7
12. Greeno, J.G., van de Sande, C.: Perspectival understanding of conceptions and conceptual growth in interaction. Educ. Psychol. **42**(1), 9–23 (2007). https://doi.org/10.1080/00461520709336915, Annual Meeting of the American-Educational-Research-Association
13. Hawkins, D.: Critical barriers to science learning. Outlook **29**, 3–23 (1978)
14. Hayashi, Y., Koedinger, K.: What are you talking about?: a cognitive task analysis of how specificity in communication facilitates shared perspective in a confusing collaboration task. In: Proceedings of the 41st Annual Conference of the Cognitive Science Society(CogSci2019), pp. 1887–1893. Cognitive Science Society (2019)
15. Hayashi, Y.: The power of a "maverick" in collaborative problem solving: an experimental investigation of individual perspective-taking within a group. Cogn. Sci. **42**(1, SI), 69–104 (2018). https://doi.org/10.1111/cogs.12587
16. Hayashi, Y., Miwa, K.: Prior experience and communication media in establishing common ground during collaboration. In: Proceedings of the Annual Meeting of the Cognitive Science Society(CogSci2009). Cognitive Science Society (2009)

17. Hayashi, Y., Shimojo, S.: Modeling perspective taking and knowledge use in collaborative explanation: investigation by laboratory experiment and computer simulation using act-r. In: Proceedings of the 23rd International Conference on Artificial Intelligence in Education (AIED2022), pp. 647–652. International Artificial Intelligence in Education Society (2022)
18. Keysar, B., Barr, D., Balin, J., Brauner, J.: Taking perspective in conversation: the role of mutual knowledge in comprehension. Psychol. Sci. **11**(1), 32–38 (2000). https://doi.org/10.1111/1467-9280.00211
19. Koedinger, K.R., Corbett, A.T., Perfetti, C.: The knowledge-learning-instruction framework: bridging the science-practice chasm to enhance robust student learning. Cogn. Sci. **36**(5), 757–798 (2012). https://doi.org/10.1111/j.1551-6709.2012.01245.x
20. Lombrozo, T.: The structure and function of explanations. Trends Cogn. Sci. **10**(10), 464–470 (2006). https://doi.org/10.1016/j.tics.2006.08.004
21. Miyake, N.: Constructive interaction and the iterative process of understanding. Cogn. Sci. **10**(2), 151–177 (1986). https://doi.org/10.1016/S0364-0213(86)80002-7
22. Richardson, D.C., Dale, R.: Looking to understand: the coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. Cogn. Sci. **29**(6), 1045–1060 (2005). https://doi.org/10.1207/s15516709cog0000_29
23. Scardamalia, M., Bereiter, C.: Computer support for knowledge-building communities. J. Learn. Sci. **3**(3), 265–283 (1994). http://www.jstor.org/stable/1466822
24. Scardamalia, M., Bereiter, C.: Knowledge building theory, pedagogy, and technology. In: Sawyer, R. (ed.) Cambridge handbook of the learning sciences, pp. 97–115. Cambridge Handbooks in Psychology, Cambridge University Press, Cambridge (2006)
25. Schmidt, R.C., Fitzpatrick, P.: The origin of the ideas of interpersonal synchrony and synergies. In: Passos, P., Davids, K., Chow, J. (eds.) Interpersonal Coordination and Performance in Social Systems, pp. 17–31. Routledge (2016)
26. Schwartz, D.: The emergence of abstract representations in dyad problem solving. J. Learn. Sci. **4**(3), 321–354 (1995). https://doi.org/10.1207/s15327809jls0403_3, 4th Annual Winter Text Conference
27. Smith, J.P., diSessa, A.A., Roschelle, J.: Misconceptions reconceived: a constructivist analysis of knowledge in transition. J. Learn. Sci. **3**(2), 115–163 (1993). https://doi.org/10.1207/s15327809jls0302_1
28. Thomson, R., Pyke, A., Trafton, J.G., Hiatt, L.M.: An account of associative learning in memory recall. In: the 37th Annual Conference of the Cognitive Science Society. Cognitive Science Society (2015)
29. Weinberger, A., Fischer, F.: A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. Comput. Educ. **46**(1), 71–95 (2006). https://doi.org/10.1016/j.compedu.2005.04.003

Does Volunteer Engagement Pay Off? An Analysis of User Participation in Online Citizen Science Projects

Simon Krukowski[1](✉), Ishari Amarasinghe[2],
Nicolás Felipe Gutiérrez-Páez[2], and H. Ulrich Hoppe[1]

[1] RIAS Institute, Duisburg, Germany
{sk,uh}@rias-institute.de
[2] ICT Department, Universitat Pompeu Fabra, Barcelona, Spain
{ishari.amarasinghe,nicolas.gutierrez}@upf.edu

Abstract. An increasing number of Citizen Science (CS) projects rely on digital technologies for data collection and processing as well as for the coordination and communication between participants. However, these projects raise the question about whether the interplay between professional scientists and volunteers actually constitutes a collaborative working relationship in terms of shared goals, initiative and responsibility. A related question is about added values and benefits that volunteers gain by their participation. These issues have been studied with a sample of projects taken from the Zooniverse platform, particularly by analysing the communications found in the projects' forum pages. Social network analysis techniques are used to identify structural characteristics underlying the forum interactions as basic measures that are repeated over time sequences to capture the dynamic changes. The results show that a smaller group of volunteer users is responsible for a large portion of communication and coordination actions, with some of them being promoted to moderators, which can be seen as a reward and incentive.

Keywords: Citizen science · Volunteer engagement · Network measures of collaboration

1 Introduction

Many current activities and projects labeled as Citizen Science (CS) rely heavily on digital technologies for collecting, storing, exchanging and processing data. However, digital communication and collaboration technologies such as forums, wikis, talk pages and other types of social media also play an important role in this respect. These types of activities have been specifically subsumed under the notion of "online Citizen Science" [16]. On the human side, professional scientists work together with volunteers to advance specific lines of research and generate new scientific knowledge. In these joint activities, volunteers often perform time-consuming data collection and analysis tasks but also take a more or less active

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 67–82, 2022.
https://doi.org/10.1007/978-3-031-20218-6_5

part in the communities formed around topics of interest. For the volunteers, this provides opportunities to enhance their own scientific skills and knowledge by participating in processes of scientific inquiry. Such collaborations between scientists and volunteers can be seen as an essential and crucial aspect for the advancement of science and also for the personal enrichment and growth of the volunteers involved. However, we still need to better understand and reflect the relationship between professional scientists and volunteers in this context. If we characterize it as "collaborative" in a strict sense, this would imply shared common goals within a jointly conceived task space. These questions have been taken up in the ongoing European research project CS Track[1] that provides the background and context of the work described here. While an earlier report on this by Amarasinghe et al. [1] was based on a single example project, for this work, we extended the scope of our analysis including the sampling and selection of projects, included new types of data (namely role changes of the volunteer participants) and refined and extended our inventory of analytics methods.

A large amount of CS activities occurs on online CS platforms such as Zooniverse[2], where volunteers mainly contribute by classifying or tagging data provided by scientists [4]. When considering the working relationship between scientists and volunteers, these projects can be characterised as *contributory* projects, as volunteers mainly contribute data (e.g., by classifying pictures of galaxies) and do not collaborate with the scientists to the extent of adjusting or formulating the research focus [4]. Although contributory projects allow for an efficient collection of data for completing the task which would otherwise be difficult to achieve (*human computation*, see [12,13]), volunteer participation in these projects is not necessarily collaborative. With less collaboration in terms of shared initiative and responsibility, the benefits and incentives for volunteers to participate become less clear, and high attrition rates threaten the sustainability and continuity of participation [2,5]. Recent studies of volunteer participation in contributory projects have led to the identification of different participation and engagement levels [8,13], yet do not provide specific insights related to the collaboration among volunteers and between volunteers and professional scientists.

The Zooniverse platform offers users the possibility to interact with each other through an integrated discussion forum. Such a forum serves as a platform for collaboration between users and different user groups and it can provide additional benefits for volunteers in terms of getting promoted to higher roles such as "moderator". In the work presented here, we focus on these discussion forums. Methodologically, we rely on social network analysis (SNA) to examine the interactions and discourse structure of volunteers in the discussion forums of CS projects hosted on the Zooniverse platform. By doing so, we can detect relational structures within the data as well as interactions between different user groups [9]. To date, few studies have considered these discourse patterns of participants in contributory CS projects to infer different levels of participation, collaboration and possible benefits. Due to the high prevalence of online

[1] https://cstrack.eu/.

[2] https://zooniverse.org/.

contributory projects, insights in this regard are highly relevant and can provide clues for the design of future CS projects and for defining appropriate incentives to increase the amount of collaboration between scientists and volunteers. Our analyses are guided by the following two research questions:

RQ1: How collaborative is the interaction among volunteers and between volunteers and professional scientists in terms of mutual engagement and initiative for the given sample of projects?
RQ2: Which benefits do volunteers gain from participating in CS projects on Zooniverse?

2 Background

In the following, we will provide the relevant background for our RQs by describing key concepts as well as related work on the analysis of volunteer participation and discourse in CS projects.

2.1 Volunteer Engagement in CS Projects

Citizen Science refers to the participation of the public (i.e., non-professional scientists) in scientific activities. As such, it gains attention both as a field of practice and as a target of research [20], as it aims to bring scientific research closer to the general public, while at the same time offering a multi-faceted area of research. Regarding the latter, there is a particular interest in exploring and understanding volunteer engagement in CS projects and the collaboration between professional scientists and volunteers [20]. Based on the level of public participation in scientific research, CS projects can be divided into three categories, namely *contributory*, *collaborative* and *co-created* projects [4]. Contributory projects are designed by professional scientists, and volunteers mainly contribute by collecting data across a wide geographic area or by classifying and annotating data they are provided with. Collaborative projects on the other hand go beyond this and also involve volunteers in other steps of the scientific process such as the refinement of the project design, analysis of data and the dissemination of findings. Co-created projects (as the name suggests) are based on joint initiatives by both the public and scientists together. Co-created projects mostly address specific concerns of the public, e.g., a problem related to the weather or environment of a local community.

Pandya has shown that active participation of citizens in every step of the scientific process is a key to a project's success [11]. According to the above mentioned categories, the CS projects hosted on the Zooniverse platform can be categorised as contributory projects, since volunteers merely classify and annotate the data they are being provided with by professional scientists and do not take part in other aspects of the scientific process. Especially in such projects, difficulties arise when scientists try to sustain continuous volunteer participation and engagement [2], and previous studies have shown that volunteers indeed engage in CS projects to varying degrees. For instance, Ponciano et al. [14] suggested two different engagement profiles (i.e. transient and regular) and have shown that volunteers who continuously contribute to the project constitute a minority. Eveleigh et al. [5] proposed that the participants can be divided into two categories namely high and low contributors based on their quantities of project contributions. In addition, prior studies have also provided evidence that high attrition rates and dabbling behaviour are threats to the sustainability of online communities and retaining participation [5]. Hence, the exploration of volunteer motivation and their degrees of engagement/disengagement could help to derive incentive mechanisms, new strategies to enhance engagement as well as implications for the design of online communities that could attract and sustain citizens' participation.

Nevertheless, existing research has mainly focused on using survey data and interviews [15,17] to understand such motivational factors related to citizen participation in CS projects, yet the tracing of behavioural patterns and the analysis of discourse data collected from online CS forums could provide additional, data-driven and fine-grained insights which help to better understand the extent of volunteer participation [3,8].

2.2 Discourse Data and Social Network Analysis (SNA)

Project forums aim to provide a space to discuss and socialise for volunteers and scientists in online CS projects. Using forums, volunteers can post questions, discuss their doubts about data collection and analysis, propose ideas and engage in a discussion with scientists to draw conclusions [3]. Functional role assignments (e.g., scientists, moderators, volunteer-moderators) are a common practice in these forums and aim to facilitate the regulation of discourse within those spaces [16]. For example, experienced users get promoted to advanced roles such as volunteer-moderators who would then bring the attention of scientists about questions which are difficult to solve using community knowledge only, or about potential findings [19]. Various techniques can be used to model discourse among different stakeholders in CS projects, for instance, by developing coding schemes, counting word occurrences or using epistemic network analysis [1]. In a particularly interesting and related approach, Rohden et al. examined how different user groups organise knowledge by using the various technical features offered by the Zooniverse discussion forums (e.g., linking, mentioning or using hashtags), finding different discourse patterns among the user groups [16]. In their case study, they could show, that the discussions are mainly dominated by highly active volunteers, scientists and moderators, and that these user groups participate in the discourse differently. For example, moderators and highly active volunteers were shown to mostly create new discussions, while scientists are often prompted to provide contextual information (e.g., by mentioning). The findings show, that the discussion forums are an important aspect of the project, which not only acts as a platform for knowledge organisation and collaboration between volunteers and scientists, but also as an incentive mechanism for certain users to participate in scientific discourse.

In this study, we extend these findings and examine them through the lens of social network analysis (SNA, see [21]). It is particularly well-suited for the analysis of discourse and relational patterns, as it can reveal relational information and developments which would otherwise not be easily observable [18]. We can do this by constructing a graph where CS forum users represent nodes, and the edges between them reflect some type of connection depending on the operationalisation of the network. In the case of forum data like on Zooniverse, such a connection can represent the act of replying to another user or commenting on a forum post. This allows us to approach **RQ1** by considering how such communicative actions relate to the productivity within the project, and what we can learn from different SNA measures of the networks. Additionally, we can move beyond the global perspective of the network and examine measures like the degree centrality of individual users to model their relative importance and see how it develops over time, providing insights with regard to **RQ2**. The SNA measures used for this will be explained in Sect. 3.3.

3 Method

For this study, we deployed several data collection and extraction steps which will now be explained. To this end, we will also describe the structure of the data and the creation of the network.

3.1 Data Collection and Sampling

We extracted data from the discussion forum known as the "Talk page" of Zooniverse CS projects, where comments are grouped by different boards. Common boards are the *Notes* board (discussion of individual subjects to be classified); *Help* (volunteers can seek for expertise and report issues); *Chat* (discussions beyond the research project, e.g., about common interests associated to the research field), and *Science* (discussions regarding the theoretical background, methods and practices around the project's research field). While more boards exist (i.e., *Announcements*), we focused our analysis on the four mentioned boards since their goal is to foster interactions between volunteers and scientists. To extract the forum data in December 2021, we used the public API services provided by the Zooniverse platform. We extracted comments of all the projects from discussions with at least 2 or more comments (a total of 2,049,646 comments from 703,139 discussions). Then, we restricted the sample to projects that used the four boards to be studied and to discussions with 10 or more comments in each of these boards. Table 1 presents the list of the seven resulting projects and the comments per board. In total, the original sample contains 24,734 comments. As can be seen, the seven projects in the original sample differ drastically with regard to their amount of interactions, as some projects only have a small amount of comments. In consequence, we chose to only consider the projects *Galaxy Zoo*, *Gravity Spy* and *Snapshot Wisconsin* for further analyses.

The comments serve as the basis of our data set. For each comment, we have information on its author, creation time, type (e.g., reply) and the role

Table 1. Number of comments per board for each of the seven projects

Project	Comments			
	Chat	Help	Notes	Science
Seabirdwatch	158	119	98	27
Galaxy Zoo	**1,156**	**1,236**	**3,498**	**4,348**
Gravity Spy	**996**	**804**	**6,573**	**694**
Snapshot Wisconsin	**551**	**334**	**2,534**	**73**
Wildwatch Kenya	190	292	397	15
Galaxy Nurseries	41	18	14	14
Penguin Watch	54	135	246	119

[1] projects in **bold** were chosen for further analysis

of its author. In the forum, the user roles visually appear as a badge in the user profiles, and the user rights depend on them (e.g., the possibility to moderate discussions). The default role without any rights is the *volunteer* role. Further roles are *team*, *scientist*, *moderator*, and *admin*. Forum users can also have multiple roles, in these cases we defined absorption rules to absorb the roles to the most relevant one while additionally considering the user history (e.g., *moderator* & *admin* becomes *moderator*). We defined the rules in such a way that we can particularly distinguish scientists from non-scientists. By considering the user history, we can also detect role changes (e.g., *volunteers* can become *volunteer-moderators*). Possible roles are *volunteer*, *volunteer-moderator*, *scientist*, *scientist-moderator* and *moderator*.

Additionally, we extracted the participation activity per project (i.e., the number of daily contributions done by the users of each of the projects) to better understand the behaviour of users in the forum in relation to the productivity in terms of contributions, resulting in the number of classifications per day.

3.2 Network Extraction

From the comment data described above, we extracted a directed network with parallel edges allowed using the NetworkX python package [7]. In the network, each user who appears in the data by having created a comment is reflected as a unique node u. For the creation of edges $(u,\ v)$, two types of possibilities were considered: **Reply** (if u replies to v) and **direct comment** (if u posts a comment in a discussion created by v). For each node u and edge $(u,\ v)$, we have several attributes from the comment data. As node attributes, we consider the role the user has, as well as topological features which can be calculated (i.e., degree centrality). For each edge, we have attributes regarding the time of creation, board title, type of relation and discussion title. As such, it is important to note that each edge corresponds to a comment which was posted (either by replying or by directly commenting). This is also reflected by the direction of the edges: Users, whose posts get commented or replied to a lot therefore have a high in-degree, for example. Due to the construction of the network, not all comments from the original sample were considered, e.g., original posts, as they do not contain any relational data. However, such users would still appear in the network as they are the target of other edges. In total, the created networks for our three analysed projects across the whole analysed time period consist of 1,490 nodes and 15,975 edges. For certain analyses, we defined time slices, reflecting a quarter of a year. We then first filtered the comments according to the time frame and extracted a new network from the filtered comments, resulting in a distinct network for each time slice. This allows us to additionally consider temporal aspects (e.g., development of avg. degree). Although there is a possibility of losing the context of discussions spanning across these time slices, this proved to be rare as most discussions occurred within our defined slices.

3.3 Network Measures

We considered several network measures to examine the communication patterns and discourse in the forums:

Degree Centrality. The most straightforward way to measure the influence and importance of a node within a network is the *degree centrality*, which simply counts the number of neighbouring nodes adjacent to a given node [21]. In directed networks, one can additionally differentiate between incoming and outgoing edges, and thus the *in-* and *out-degree*. Nodes with higher degree centrality are considered to be more prominent and influential.

Average Degree. The average of the degree values for the whole network is an indicator of overall connectivity, which is less prone to scaling effects than the *density* measure which tends to decrease in growing networks.

Centralisation. The *centralisation* of a graph reflects the inequality of degrees between the most central node and all the other nodes [6]. A high value of centralisation indicates a hierarchical structure of a graph whereas low values are characteristic for a more homogeneous distribution of connections. E.g., a "star" graph with one node in the center connected to all others but no interconnections between these other nodes would maximize the centralisation. In general, higher values of C_D indicate more inequality between the nodes regarding their centrality, which can be an indicator for less collaboration among all forum users. Equation 1 shows how the centralisation C is calculated from the degree values C_D of all nodes as follows (p_i running through all nodes and p^* representing a node with maximum degree):

$$C = \frac{\sum_{i=1}^{n}[C_D(p^*) - C_D(p_i)]}{max \sum_{i=1}^{n}[C_D(p^*) - C_D(p_i)]} = \frac{\sum_{i=1}^{n}[C_D(p^*) - C_D(p_i)]}{(n-1)(n-2)} \tag{1}$$

Reciprocity. Another global network measure that captures a characteristic of collaborativeness is reciprocity R [21]. It describes the ratio of edges pointing in both directions to the total number of edges in the graph. Thus, if there are many mutual, bi-directional edges in the network, the reciprocity R is high and this would be a positive indicator for collaboration.

4 Results

To answer our RQs, we conducted several analyses which will now be described. We will first describe our analyses for **RQ1** on the relation of collaboration and productivity as well as the distribution of activity per user role. This will be followed by our analyses regarding **RQ2** where we examined the participation of certain highly active users who changed their role over the course of the project.

4.1 Interactions over Time

We considered the edges of the network for communication and the number of contributions per day for productivity.

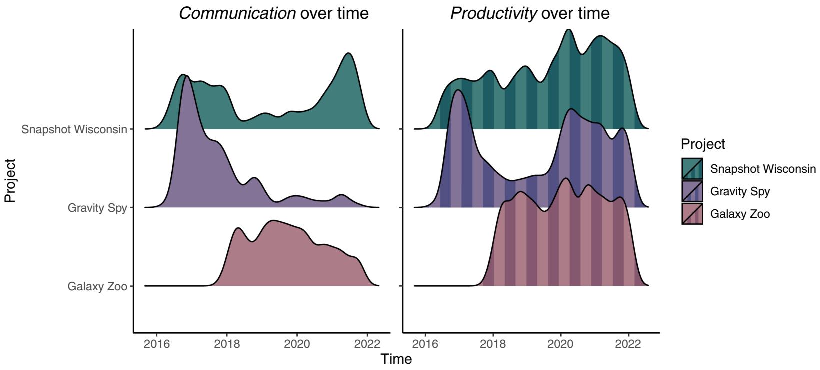


Fig. 1. Density diagram for collaboration (edges) and productivity (classifications) for the three analysed projects across time

As can be seen in Fig. 1, certain trends show for the projects. While communication seems to decrease in *Gravity Spy* and *Galaxy Zoo*, it appears to increase in *Snapshot Wisconsin* towards the end. Especially during the beginning of the projects, communication and productivity seem to correlate, while they do not necessarily do so during later phases. Statistically, for *Galaxy Zoo* and *Snapshot Wisconsin*, we did not observe any significant correlation (Pearson's r = .37 and r = .35, respectively), however for *Gravity Spy*, communication and productivity appear to correlate moderately (r = .48, p = .02). For *Snapshot Wisconsin* and *Gravity Spy*, productivity increases again towards the end of the examined time frame, which is accompanied by an increase in collaboration only for *Snapshot Wisconsin*. Interestingly, we see a peak in productivity for all projects around 2020, which can likely be attributed to the pandemic. Thus, communication through the forum can be observed across the whole time span and is particularly high around the beginning of the project.

4.2 Centrality over Time

In a next step, we considered the different network measures described in Sect. 3.3 and how they behave over time, as insights in this regard are particularly relevant for **RQ1**. Figure 2 shows these measures over the observed time span in respect to the productivity. High values of avg. degree and reciprocity and low levels of centralisation hint at more collaboration. As a general observation, productivity appears to go along with the collaboration measures across our three analysed projects. Despite some fluctuations, the avg. degree increases over time for *Galaxy Zoo* and *Snapshot Wisconsin*, and decreases for *Gravity Spy*. Except for *Gravity Spy*, the reciprocity goes along with productivity, meaning that more mutual discussions occur when productivity is high. Interestingly however, in *Gravity Spy*, there are many fluctuations in reciprocity: Particularly at times when there is high productivity, reciprocity is low, and similarly vice-versa. Centralisation increases for *Galaxy Zoo* and *Snapshot Wisconsin*, while it generally appears to decrease for *Gravity Spy* after peaking around 2019. To infer the collaboration however, we have to consider the interplay of these measures. High centralisation paired with a high avg. degree during times of low productivity in *Gravity Spy* for example indicate that only few users kept participating in the project, yet heavily influenced most of the discourse. To this end, *Gravity Spy* shows more signs of collaboration during the first half of the project, as avg. degree and reciprocity are high, yet the centralisation is low. Similarly, *Snapshot Wisconsin* shows more collaboration around 2020, while no clear picture emerges for *Galaxy Zoo*. Generally, it can be said that despite these

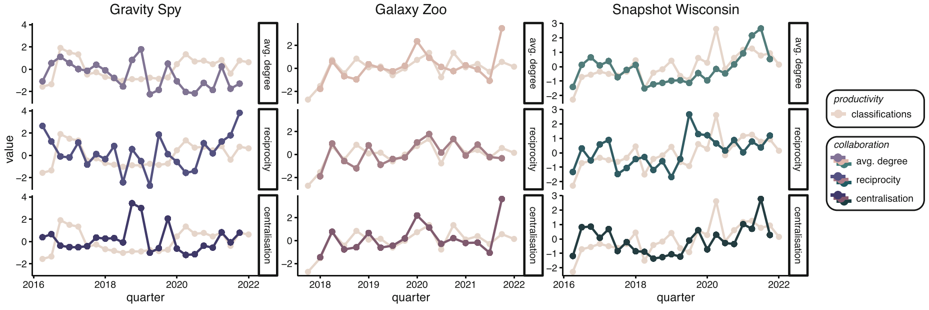


Fig. 2. Collaboration measures across time for the three examined projects. The values are z-standardised to allow for comparability

fluctuations, the projects appear to be collaborative, as we cannot observe that collaboration and productivity drift apart (i.e., only correlate in the beginning when volunteers need help). However, the high centralisation paired with a high avg. degree for *Galaxy Zoo* and *Snapshot Wisconsin* hints at certain users who are highly active.

4.3 Participation by User Role

Of particular interest for our **RQ1** and the collaboration between volunteers and scientists is the consideration of the different user roles (see Sect. 2.2) which users are assigned to when participating in the forum. As can be seen in Fig. 3, *volunteers* account for the majority of interactions (i.e., edges), with the exception of *Snapshot Wisconsin* where *volunteer-moderators* and *scientists* also shape the discourse to a large extent.

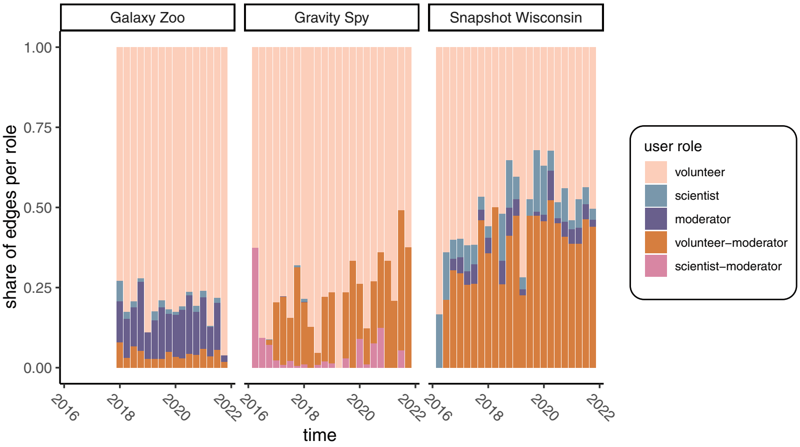


Fig. 3. Relative participation by user role

Interestingly, *scientists*, *moderators*, and *volunteer-moderators* all appear to contribute early on in the project, but this contribution diminishes over time. An interpretation for this decrease could be that over time, *volunteers* need less help, as they get familiar with the task. Therefore, *scientists* and *moderators* need to give less help in the forum, which is reflected by a lower participation. An exception to this is the increase in participation by *volunteer-moderators* over time, which is particularly true for *Gravity Spy* and *Snapshot Wisconsin*. Within these projects, *volunteer-moderators* continually increase their share in the discussions, nearly accounting for 50% of the comments made. As described in Sect. 3.1, *volunteer-moderators* are users who were initially volunteers and then obtained moderator rights. Their high share in the discussions is of special interest, specifically with respect to **RQ2**, as it is expected that such users are highly motivated as the promotion to higher user roles might serve as an additional incentive to participate in the forum.

4.4 Participation and Rewards

Our analyses showed that certain users changed their roles over the course of the project. For our three analysed projects, we observed 14 role changes, of which 3 were excluded (due to users changing their role within a single day), resulting in a total of 11 users who changed their role (out of 1490 users in total). In 10 of those cases, users changed their role from *volunteer* to *volunteer-moderator* and thus got promoted to a higher role. In one case, a *scientist* additionally obtained the moderator role and became a *scientist-moderator*. Although the amount of role changes appears to be small, the proportion of these users in the general discussions is substantial. This can be seen in Fig. 4 for the *Gravity Spy* project, which shows the relative proportion of comments per respective time frame by users who at some point in the project changed their role. These users account for nearly 40% of all the comments made during the project's lifetime, and this proportion is dominated by a few single users. For example, *user2* is responsible for a substantial proportion of the comments during the first half of the examined time span, and is then superseded by *user6* towards the end. To extend this finding, we compared the average degree between users who changed their role and those who did not. As a criterion, we excluded users without substantial participation and only included those with a degree >30, resulting in two groups of differing sizes, but comparable standard deviations (see Table 2).

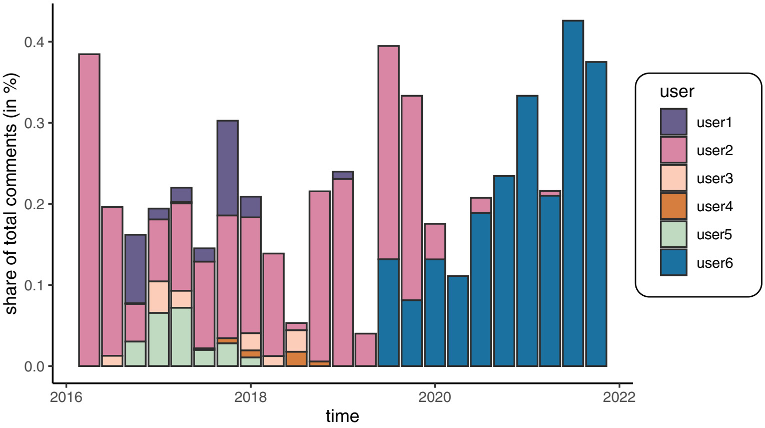


Fig. 4. Relative share in discourse by users with role changes for the *Gravity Spy* project

Table 2. Collaboration of role change users vs. non-role change users

Role change	N	Degree		In-degree		Out-degree	
		M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
True	10	464.80**	258.78	213.60**	159.59	251.20**	182.52
False	135	134.41**	254.41	70.20**	127.81	64.21**	134.67

** significant difference with p <.001

It shows that the degree centrality drastically differs between the two groups, with those users who changed their role having a degree more than three times higher than those who did not change their role. This also shows for the in- and out-degree, meaning they both give and receive more comments. A Wilcoxon rank sum test (due to absence of normality) showed a statistically significant difference between these means. Thus, it shows that users who changed their role over the course of the project and got promoted to a higher role appear to contribute significantly to the discussions by collaborating with other users.

4.5 User Trajectories

To better understand these highly engaged users, we examined their individual development across the projects. To this end, we considered three different indicators, specifically their degree, the ratio of in- vs. out-degree and their periodicity. For the degree, we calculated ranks in order to account for fluctuations in the general participation within the projects. Lower ranks indicate higher degrees (i.e., the first rank meaning the user had the highest degree during this respective time frame). The in- and out-degree refers to the direction of the edges, as explained in Sect. 3.3. There, shifts can tell important insights regarding the importance of the user in the network. Lastly, we considered the periodicity of the user, which was calculated similarly to [13], by considering the average days of absence (i.e., no comments made) as well as the standard deviation of these days of absence per time frame. As such, low values for the mean indicate high adherence. We calculated these measures for the 11 users who changed their roles, and when applying the same filter as in Table 2, it shows that on average, the degree rank is 13.29 (SD = 6.07) for users without a role change (N = 135), and 5.03 (SD = 2.18) for users with a role change (N = 10). A similar picture emerges regarding in- and out-degrees, with 8.91 (SD = 9.78) and 9.30 (SD = 12.70) for users without a role change and 14.40 (SD = 13.42) and 14.95 (SD = 9.15) for users with a role change. Due to missing data, no mean values for periodicity could be calculated. Thus, the analyses again reveal that users who changed their role appear to be highly connected and consistently participate in the discussions. Figure 5 shows the trajectory of a user in the *Snapshot Wisconsin* project with a role change in the beginning of the "career" who additionally obtained moderator rights by getting promoted to *volunteer-moderator*. This user occupied ranks between 4 and 6 related to degree before the role change and then consistently stayed in the top three ranks afterwards. This is also reflected in a

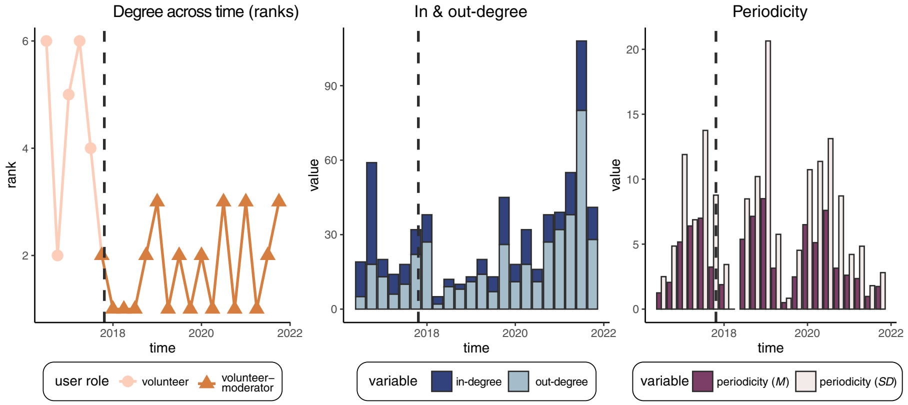


Fig. 5. Trajectory of a user with a role change from *volunteer* to *volunteer-moderator*. The dashed vertical line indicates the time of the role change

steady increase of the absolute degree values. Thus, over time, the user becomes more central in the network. Particularly relevant is the distinction of in- vs. out-degree, as it can be seen that their ratio inverted after the role change. Before the role change, the user received more comments/replies, but afterwards, they gave more comments/replies. Around the time of the role change, the user also stayed on the platform most consistently, as reflected by the low periodicity. A similar picture with low periodicity and lower degree ranks around or after the time of the role change emerges for the majority of the other examined users with role changes.

5 Discussion

We examined the participation of volunteers in contributory CS projects, specifically using data records of forum interaction to assess how collaborative the interactions in a given project are (**RQ1**) and which benefits volunteers can gain from their participation (**RQ2**). We approached these research questions using SNA to analyse user interactions extracted from Zooniverse forum data for three selected projects. Our analyses confirm related studies on this topic [10] by showing that the analysed projects appear to be mostly collaborative (**RQ1**) as volunteers use the forum to help each other and exchange knowledge, as shown by the correlation between communication and productivity (i.e., classifications). Especially at the beginning of the projects, communication appears to be high. This could be due to several reasons such as volunteers who introduce themselves or regarding the project task, as volunteers are not yet familiar with the topic and seek for help. The peak in interactions around 2020 can be attributed to the pandemic, as it can be assumed that many users spent their time during lockdowns participating in these projects. Our analysis of user roles confirmed the need for help in the starting phase, as the share of moderators and scientists in the comments is particularly high in the beginning. The forum then acts as a collaborative space for knowledge exchange. The application of the network measures indicating collaboration introduced in Sect. 4.2 further proved this, showing that, on average, users become more central over time in *Galaxy Zoo* and *Snapshot Wisconsin*. However, the network also becomes more centralised, indicating a concentration of few users. Particularly interesting is the high centralisation and low reciprocity during the gap in productivity for *Gravity Spy*. This corresponds to a kind of emergency situation in which only few highly central users keep the activity going without much active participation from others.

With regard to **RQ2**, our analyses show that although only a small fraction of users change their status and get promoted to a higher role, these are responsible for a substantial portion of the comments and are three times better connected than users without a role change. Additionally, they steadily increase their share in the discourse. While this confirms recent studies showing that a minority of users is responsible for the majority of the discourse in CS forums [10,16], our results add the connection to role changes. Such role changes require certain users to be willing and motivated to assume additional rights and responsibilities in the forum interactions. Such a promotion can be seen as a reward and personal benefit for the volunteers. A closer look at the individual trajectories of these highly engaged users additionally showed that the role change is also visible in their participation patterns: After the role change, these users consistently stay in the top ranks with regard to their centrality, their commenting behaviour changes (i.e., ratio in- vs. out-degree) and they participate in the forum very frequently without many days of absence. Thus, user roles, and especially the change of them, seem to play a significant role in the participation of volunteers in CS projects, and might serve as an intrinsic benefit, setting these projects apart from mere crowdsourcing projects with aspects of human computation. As such, the findings inform design implications for new online CS projects as they suggest that role changes serve as an incentive to continuous and collaborative participation of volunteers in these projects. However, the findings of the present study should be interpreted in light of the following limitations. Although we carefully chose the projects for our sample based on the criteria explained in Sect. 3.1, the data might still be biased and look differently with another choice of projects, and our computational limitation to only choose discussions with at least 10 comments might have also influenced the results. Moreover, the types of classification tasks proposed, socioeconomic backgrounds and ages [8] may have influenced the observed participation patterns of volunteers. Therefore, these aspects require further research, preferably by deploying our analyses on bigger samples which involve more projects.

6 Conclusions and Future Work

As compared to our prior work [1], we have extended the sample of projects analysed and refined our analytics approach, especially by examining time series of networks and including information on the user status in terms of role changes. Our sample is still homogeneous in that it only contains projects from one online CS platform, namely Zooniverse. All projects on this platform are considered as "contributory" and request data processing and annotation tasks to the volunteers, yet not the collection of source data. This is a quite instrumental task definition, yet we still see that a large of the interactions are maintained by a small number of active users with a high share of volunteers among these. The presence of data collection activities might lead to a higher sense of ownership regarding the project work and outcome and should thus even increase participation. Yet, even for the task profile specific to Zooniverse we could see the relevance of volunteers in the overall discourse manifested in the forum as a space of coordination and exchange. We could also see that although not all highly active volunteers are promoted to moderators, those who are promoted show significantly higher engagement in the interactions. At least for these volunteers the engagement is rewarded or "pays off". Network-analytic participation measures allow for identifying highly engaged users and may be used as a basis for making better informed decisions about these types of rewards.

Acknowledgements. This work was partially funded by the European Union (EU) in the context of the CS Track project (Grant Agreement no. 872522) under the Horizon 2020 program. This document does not represent the opinion of the EU, and the EU is not responsible for any use that might be made of its content. We thank all CS Track team members for the fruitful collaboration.

References

1. Amarasinghe, I., Manske, S., Hoppe, H.U., Santos, P., Hernández-Leo, D.: Using network analysis to characterize participation and interaction in a citizen science online community. In: Hernández-Leo, D., Hishiyama, R., Zurita, G., Weyers, B., Nolte, A., Ogata, H. (eds.) CollabTech 2021. LNCS, vol. 12856, pp. 67–82. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-85071-5_5
2. Aristeidou, M., Scanlon, E., Sharples, M.: A design-based study of citizen inquiry for geology. In: Hernández-Leo, D., Ley, T., Klamma, R., Harrer, A. (eds.) Eighth European Conference on Technology Enhanced Learning - EC-TEL 2013. LNCS, vol. 8095, pp. 7–13. Springer, Cham (2013). https://doi.org/10.13140/RG.2.1.1265.4324
3. Aristeidou, M., Scanlon, E., Sharples, M.: Profiles of engagement in online communities of citizen science participation. Comput. Hum. Behav. **74**, 246–256 (2017)
4. Bonney, R., et al.: Public participation in scientific research: defining the field and assessing its potential for informal science education. A CAISE inquiry group report. Technical Report (2009). https://eric.ed.gov/?id=ED519688
5. Eveleigh, A., Jennett, C., Blandford, A., Brohan, P., Cox, A.L.: Designing for dabblers and deterring drop-outs in citizen science. In: Conference on Human Factors in Computing Systems - SIGCHI 2014, pp. 2985–2994. ACM, New York (2014)
6. Freeman, L.C.: Centrality in social networks conceptual clarification. Soc. Netw. **1**(3), 215–239 (1978)

7. Hagberg, A.A., Schult, D.A., Swart, P.J.: Exploring Network Structure, Dynamics, and Function using NetworkX. In: Varoquaux, G., Vaught, T., Millman, J. (eds.) Proceedings of the 7th Python in Science Conference, pp. 11–15. Pasadena, CA USA (2008)
8. Herodotou, C., Aristeidou, M., Miller, G., Ballard, H., Robinson, L.: What do we know about young volunteers? an exploratory study of participation in zooniverse. Citizen Sci.: Theor. Pract. **5**(1), 2 (2020)
9. Hoppe, H.U., Harrer, A., Göhnert, T., Hecking, T.: Applying network models and network analysis techniques to the study of online communities. In: Cress, U., Moskaliuk, J., Jeong, H. (eds.) Mass Collaboration and Education, pp. 347–366. Springer International Publishing, Cham (2016). https://doi.org/10.1007/978-3-319-13536-6_17
10. Luczak-Rösch, M., Tinati, R., Simperl, E., Van Kleek, M., Shadbolt, N., Simpson, R.: Why won't aliens talk to us? Content and community dynamics in online citizen science (2014). https://eprints.soton.ac.uk/363523/
11. Pandya, R.E.: A framework for engaging diverse communities in citizen science in the US. Front. Ecol. Environ. **10**(6), 314–317 (2012)
12. Peplow, M.: Citizen science lures gamers into Sweden's human protein atlas. Nat. Biotechnol. **34**(5), 452–453 (2016)
13. Ponciano, L., Brasileiro, F.: Finding volunteers' engagement profiles in human computation for citizen science projects. Hum. Comput. **1**(2) (2014)
14. Ponciano, L., Brasileiro, F., Simpson, R., Smith, A.: Volunteers' engagement in human computation for astronomy projects. Comput. Sci. Eng. **16**(6), 52–59 (2014)
15. Reed, J., Raddick, M.J., Lardner, A., Carney, K.: An exploratory factor analysis of motivations for participating in zooniverse, a collection of virtual citizen science projects. In: Hawaii International Conference on System Sciences - HICSS 2013, pp. 610–619. IEEE (2013)
16. Rohden, F., Kullenberg, C., Hagen, N., Kasperowski, D.: Tagging, pinging and linking - user roles in virtual citizen science forums. Citizen Sci.: Theory Pract. **4**(1), 19 (2019)
17. Rotman, D., et al.: Dynamic changes in motivation in collaborative citizen-science projects. In: Conference on Computer Supported Cooperative Work - CSCW 2012, pp. 217–226. ACM (2012)
18. Sbrocchi, C., Pecl, G., Putten, I.v., Roetman, P.: A citizen science community of practice: relational patterns contributing to shared practice. Citizen Sci.: Theor. Pract. **7**(1), 3 (2022)
19. Tinati, R., Van Kleek, M., Simperl, E., Luczak-Rösch, M., Simpson, R., Shadbolt, N.: Designing for citizen data analysis: a cross-sectional case study of a multi-domain citizen science platform. In: 33rd Conference on Human Factors in Computing Systems - CHI 2015, pp. 4069–4078. ACM (2015)
20. Vohland, K., et al.: The Science of Citizen Science. Springer Nature, Heidelberg (2021)
21. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press, New York (1994)

Effects of Digital Avatar on Perceived Social Presence and Co-presence in Business Meetings Between the Managers and Their Co-workers

Mika Yasuoka[1](✉), Marko Zivko[1], Hiroshi Ishiguro[2], Yuichiro Yoshikawa[2], and Kazuki Sakai[2]

[1] Roskilde University, 4000 Roskilde, Denmark
mikaj@ruc.dk
[2] Osaka University, 560-8531 Osaka, Japan

Abstract. Due to the Covid-19 outbreak, more people in the workforce, especially in the IT industry, started working from home. This brought a set of issues and challenges for both workers and companies across the globe such as losing touch with other co-workers. This could potentially result in decrease of the performance and innovation. This paper investigates effects of using digital avatar robots in virtual meeting environment, specifically, focusing on the perception of social presence and co-presence between workers and their managers. Our experiment results showed that majority of participants felt an improvement in social presence, co-presence and overall virtual meeting experience while using digital avatar for their meetings, especially to those who has a meeting with less familiar persons or persons from the higher corporate hierarchy.

Keywords: Digital robot avatars · Video conference tools · Social presence · Co-presence · Remote work · Cooperation

1 Introduction

In January 2020, the Europe saw its first Covid-19 case in France [1] and instantly drastic changes happened. EU countries were forced to impose new measures and strict rules to fight the pandemic and stop transmission of the virus to keep their citizens safe. Covid-19 pandemic was a unique situation for todays' population and a lot of new challenges have arisen. A serious set of economic and financial problems was also coupled with another type of issue [4], which brought greatest sufferings to general population. Many faced to "limitation of direct contact with people, restrictions on movement and travel, necessary changes in activity lifestyle, boredom and monotony and uncertainty about the future" [3]. The financial, economic, and societal problems created unprecedented challenges for industries across the globe and many people in the workplaces had faced challenges that was never seen before.

In technologically advanced countries, Covid-19 made people in the workplaces, especially in the IT sector, change their working location to their private home immediately. During this period, the steep increase was seen in the usage of video conferencing

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 83–97, 2022.
https://doi.org/10.1007/978-3-031-20218-6_6

platforms such as: Zoom, Microsoft Teams, and Slack [2]. Although these applications brought a lot of new features during the pandemic period and currently provide exceptionally rich function with good interface for both business and personal meeting experience, they lack some crucial components of in-person meetings and not manage to provide a touch of social presence.

Since April 2020, 37% of American workforce was working full time from home, compared to only 5% or 3 days per week before the Covid-19 pandemic [3]. Microsoft [3] reported the relationship between the workers stagnated and less interconnected under the pandemic period. This was not a surprise, as interactions with co-workers were normally limited to video conferencing platforms. There are other solutions than well-known video platforms such as using digital robot avatars, which could potentially improve the overall satisfaction and online meeting experience among co-workers. During heavy travel restrictions, digital robot avatars have helped customer relations in a unique way by providing novel solutions for engaging the customers and potential new clients. In addition, virtual meetings with important or potentially new clients could be especially challenging because two parties might not know each other well. Establishing trust and respect is crucial for closing a successful business arrangement [8], but it could be known as exceedingly difficult by using the video conferencing tools.

This paper, by focusing on digital avatar, emerging choice of online conference, research on whether digital avatar can be used to surpass some of the limitations that are noticed in well-known video conferencing platforms. This paper focuses on investigating potentials of digital robot avatar as a substitute of face-to-face meeting in a business context. More specifically, this research is conducted in a partnership with a Copenhagen office in Denmark of one of the biggest IT companies in EU [6]. The company was chosen as it has conducted many initiatives to strengthen the well-being of their workers [7], embraced working from home culture, and managers are often not co-located even before the pandemic. The employee working in the Copenhagen office typically work with their managers located outside of Denmark such as Norway, Ireland, United Stated and Spain. This specific setup creates a unique situation in which both sides were very dependent on the video conferencing platform and cannot be efficient without them under the strict travel restrictions. By aiming at increasing the levels of satisfactions for the online meeting experiences, this research investigates how usage of digital avatar in office environments can help managers manage and cooperate with their workers in different countries, and at the same time what the impact on social presence, co-presence and overall satisfaction is while using such digital avatar in the meetings.

The rest of the paper is organized as follows. First, related literatures and works are introduced. Next, methods and experiments of the research is introduced. Finally, collected data and results are presented with concluding findings.

2 Related Work

There are some studies related to avatar robot and its perceived social presence and co-presence. In this paper, we touch upon communication, embodiment, and size of avatar robot in relation to perceived social presence and co-presence.

2.1 Communication

Social presence is important for communication and information sharing in the robot conferencing, and studied widely for a decade [9, 11]. Not many researches have been conducted in works settings but other domains such as academic conferences, education, and healthcare, social presence of robotics and business meetings have been investigated. It is well-known that interactive communication on VR among academic conference audiences can be easily achieved and provided valuable experience among participants, despite online interactive discussion have been reported as one of the biggest challenges on online academic conferences. Additionally, if the conference is conducted as hybrid conference, it has been reported that the satellite participants (online participants) were often left away from physical participants.

In the context of hybrid learning among university students, Gleason and Greenhow firmly concluded its importance in establishing meaningful bonds among students and professors, which is directly related to both the level of engagement during classes and learning outcomes [5]. The perception of trust in online communications is related important aspect [9] as establishing trust can be extremely hard if the participants do not know each other and if their only interaction and presence happens virtually in a video meeting room. In relation to trust, Jung et al. [10] articulated the importance of embodiment as “trust needs touch” [9]. Also, Brown and Prilla [13] confirmed in their AR consultancy experiment that trust was easier to establish if the avatar was by design like humans and was embodied, compared to abstract avatars. In using robot telepresence systems, many improvements should be implemented in nonverbal communication since human gestures and movements which could not be seen and transmitted in a normal video call. Another level of communication between speaking using avatar communication systems [9] should be considered.

2.2 Embodiment

Jung and Lee [12] studied many interesting concepts regarding the relation of physical embodiment to social presence while using social robots. Social robots are precisely designed for establishing social interactions with humans and can be used in different situations. Applications of social robots can be seen in areas such as education, public health care and especially elderly care and their usage is continuing to increase [12]. Still social robots necessary do not need to be physically embodied since their only task is to interact with humans on social level, unlike robots used in the manufacturing industries which must be physically embodied to move, pick, assemble and build desired products. For example, a couple of studies on robot medicate communication with advanced digital social robots have been conducted [20]. However, the importance of physical embodiment should not be neglected as Kerstin et al. [14] confirms that physically embodied social robots were perceived more attractive by the people, at the same time achieved better results on social presence perception [12].

2.3 Appearance Such as Size and Gaze

Size is also a crucial feature for achieving better results in perceived social presence and co-presence while virtually communicating with other participants. In the experiments using virtual avatar using AR platform for online consultations, Brown and Prilla [13] shows the size of online virtual consultant matter. In their research, comparing life-sized and miniature avatars, life-sized virtual avatars achieved higher levels of perceived social presence while miniature avatars were perceived as cute and pleasant. Such a characteristic of miniature avatars brings improved "perceived social presence towards the expert" [13]. This indicates that both miniature and life-sized avatars improved perceived social presence while possibility to gain trust and likeliness in a participant using miniature digital avatar was high. In the robot video conference system experiment, Rae et al. [16] confirmed the importance of size of the robots as well. In their study, as the higher setup of digital robots were, more influential on others compared to the one with lower height. Rae et al. [16] confirmed also on a correlation between possibility of avatar's ability to move and social presence and co-presence: "(T)hey preferred a dynamic avatar over a static of comparable size, 9 participants stated that they liked the more life-like movement of the dynamic avatar" [13]. Spatial awareness also proved to be one of the critical factors defining the quality and overall satisfaction of online communication [17]. Nguyen and Canny [17] researched on spatial and gaze awareness with three different gazes which heavily impacted overall satisfaction of the online meeting sessions.

The presented functionalities such as size and embodiments are all contributing to higher levels of perceived social presence and making the overall conversation make closer to a normal human conversation. Human likeness of robot or digital avatars is its resemblance to a real human being and characteristics that are intricately connected with the appearance of human being [14]. Like the physical embodiment, human-likeness of the digital avatar is also crucial while trying to establish meaningful social connection. Fong et al. [15] showed correlation between human-likeness and higher levels of social affordances, which directly made people trust and like the robot more while communicating or cooperating on certain tasks. However, it is important to note that in using the digital avatar, too much resemblance to real humans can make participants feel strange and unnatural as reached to "Uncanny Valley" effect [17].

3 Experiment

3.1 Avatar Robot Application: CommU

The tool used throughout the research was a semi-autonomous social CG-avatar room developed by one of the authors and it uses a digital avatar robot to represent participants in the conference meeting. The application allows two basic ways of setting up the video call: (1) CommU-Talk and (2) CommU-Conference. Both Talk and Conference options enable participants to communicate with each other like many other already known video platforms. What makes CommU special is digital avatars that are assigned to each participant in the call. Participants are not showing their own face and don't need to use video camera from their computer because they are represented as robotic avatar (See Fig. 1). CommU is easy to use, which only requires internet connection and a basic browser, thus, the overall setup of the system was very straightforward and easy to implement.

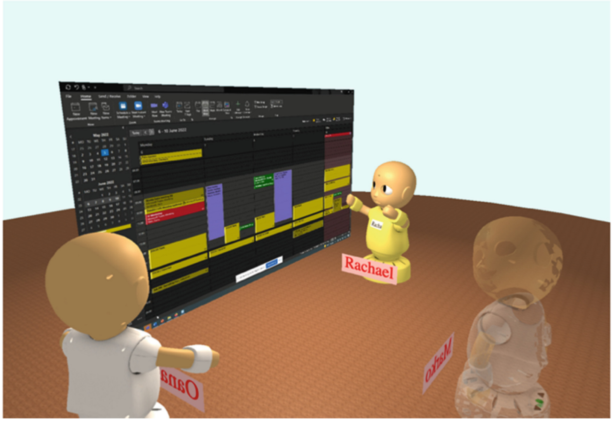


Fig. 1. Screenshot of the CommU virtual meeting session during 1–1 participant's call

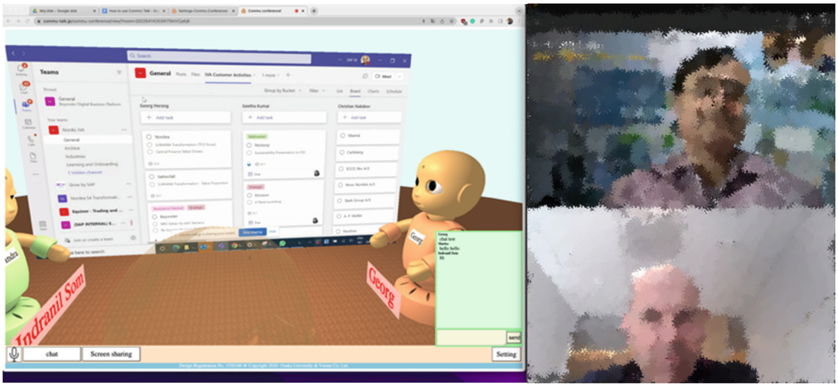


Fig. 2. By-side view of CommU Conference with chat and Microsoft Teams application

Digital avatars also implement a set of intelligent features which resemble human interaction while speaking and communication with another person. While on the call with other participants, digital avatar will move hands while speaking to signal that the voice is coming from that exact avatar and, in contrast, avatar will nod its head while listening to others when they are speaking without any users' manipulation. Also, if more than two participants are in the session, avatar will make head movements in direction of the person which is talking at that exact moment, which is another quite common gesture in human communication. Head movements of the avatar can be also controlled by clicking the mouse on certain parts of the screen where the participant would like to focus on. Depending where the person clicked with their mouse on the screen, that will be the direction where the digital avatar will turn its head. Participants also have the possibility to use the chat functionality to send and receive messages (See Fig. 2).

Concerning the overall physical appearance of the digital avatars, all are in a baby shape format. While setting up the meeting session, each participant can choose between

16 distinct types of avatars ranging from different colors, genders, and professions like: doctor, teacher, secretary, cook, office worker, etc. Depending on the situation and scenario that the participants will use CommU application, different types of avatars can be used in order to suit the environment better. The limitation in CommU-Conference, is that only three avatars with the choice of their shirt colors, compared to CommU-Talk which can chose 16 appearance. Where CommU-Conference excels compared to CommU-Talk is with the screen sharing functionality. Screen sharing is a quite common feature all well-known video conferencing tools use and especially crucial in business meetings to present essential information to other participants. CommU-Conference provides necessary key functions during business meetings. Finally, CommU application allows participants to change the settings and adjust camera angles to have a better view of the shared screen and other avatars join in the session.

3.2 Experimental Design and Participants

In this study, we conducted an experiment to investigate perceived social presence and co-presence. First, participants for this research were both full time and part time workers pf a large IT firm, ranging from student level positions to upper management positions and regional unit leads (See Table 1). Half of the participants were located in the Copenhagen office and all of those participants had their managers and supervisors located in a different office in another country like United States, Ireland, Spain, and Norway. Overall, five pairs of workers and managers participated in the research accumulating to ten people in total and nine agreed to respond to survey questions after the meeting session. Participants were all adults between the age of 24 and 52, 4 males and 5 females ranging from different nationalities, having different educational backgrounds, and working in different departments at the company ranging from: technology consultants, unit leads, finance and accounting, marketing experts and student workers. Some participants also reported that they had some experience with using digital avatar and knew in general about the importance and potential applications of digital avatars and how they could be beneficial in the future, $M = 3.4$ and $Mdn = 4.0$ *(1 = no experience; 7 = expert user in this field)*. This result was not a surprise since all the participants had extensive experience of working in the IT sector and were mostly for a few or more years in this field.

The research design was following. In the first stage each pair of workers, worker in Copenhagen office and their supervisor or manager outside Denmark would have their routine 1–1 meeting using CommU Conferencing video platform. Usually, 1–1 meetings are held once per week, lasting 30–45 min per session and are used for aligning tasks, time schedules and plans for the upcoming week and reporting on all the finished assignments during the previous week. At the company, all workers have access to a full suite of Microsoft applications and because of that, usually for all video and online meetings the company's workers use Microsoft Teams application. For research and getting relevant results, one meeting session for each pair of participants would be held primarily using CommU Conferencing tool. Before the actual research started, each participant from the Copenhagen office had a short 15-min introduction call to learn how CommU Application works and to avoid any unnecessary mistakes during the actual meeting session.

Table 1. Age and role of nine participants in the research.

	Age	Role		Age	Role
1	52	Industry value advisor	6	24	Business architect
2	51	Solution specialist	7	31	Business development
3	50	Customer innovation	8	36	Cloud solutions architect
4	47	Industry value advisory manager	9	35	Customer experience solution advisor
5	44	Business architect			

After the introduction phase with the participants from Copenhagen, CommU Conferencing virtual room was created and firstly one of the authors joined the environment in order to test whether everything worked and then other two participants were asked to join the newly created CommU Conferencing session. In total, three participants joined the conference call. Participants were given instructions to wear headphones in order to reduce the unnecessary noise and stop the echoing effect which caused a lot of problems during communication. Also, before the official start of the meeting, brief introduction to the system and functionalities were given to the other participant which did not complete introduction phase to fully understand how the application works and be able to use all the features. After the introduction phase, the official meeting started.

The participants were each assigned a robotic avatar and placed in a virtual meeting room. The third participant, who was the author, was just observing the meeting session and taking notes about interesting observations that could be useful for the research. Other two participants, who were worker in Copenhagen and manager outside of Denmark continued with their meeting using all the features that were at that moment available on CommU Conference platform, as they would usually do using Microsoft Teams application. After the 1–1 meeting session was finished, the workers from Copenhagen office stayed for a 20–30-min interview and discussion about the experience they had using CommU Conferencing platform. Because of the company's policy to flexible working, most of the interview sessions were held remotely using the company's preferred way of online communication – Microsoft Teams platform.

Lastly, after all the interviews were conducted, sessions with participants finished and results from the survey gathered final dataset was created. After the dataset was generated, it gathered all the responses coming from a survey and was later used to present meaningful results in charts and graphs using RStudio. A tool used for gathering the survey responses was Google Online Form creator which also comes with a build-in feature which generates charts and graphs and offers insights of the data collected by the application. RStudio and R language offer much more options and functionalities. Using the R language and capabilities of RStudio the author was able to present data in a visually pleasant format and at the same time extract meaningful and important calculations to present them in the conclusion. Version of the RStudio used during the research was: *2022.02.2 + 485 "Prairie Trillium"* and platform used to run operations and use RStudio application was the MacOS platform.

3.3 Data Collection

In total, nine survey responses were collected from the participants in the research process. In addition, individual interview sessions with participants from Copenhagen were collected. Thus, the data of the full online meeting sessions collected for this research originated from three main sources: 1–1 live CommU Conference video sessions with participants and observations noted during the session, individual interviews with the participants from Copenhagen office and participants' answers to a questionnaire after the meeting session and interview process was finished. Images and screen recordings were also saved by capturing the monitor and were also used to examine the recordings in greater detail. Each participant used their own computer to be able to participate in the CommU Conference call and all the participants gave their consent to use and collect data for the purpose of this research.

4 Results

The results from three main sources generated both quantitative and qualitative data. Observations gathered while actively listening to participants' meeting sessions and interviews conducted with participants after the meeting session produced qualitative data, while, in contrast, survey answers collected at the final phase as quantitative results. Nine out of ten participants agreed to answer the questions presented in the survey, which created a small but rather unique data set. It is important to note that our results are not suggested to lead any conclusions and general findings outside the area that this paper specifically discussed since the dataset is small.

Just from briefly glancing over the data it was evident that participants saw potential of the technology, digital avatar. Since participants came, as already mentioned, from the IT industry and had a very frequent occurrence of participating in online video calls, our solutions offered multiple benefits. The results shown that seven participants preferred working either fully from home or hybrid work (combining working from home with occasional visits to the office). Only two participants preferred working full time in the office.

4.1 Awareness and Knowledge About Avatar

Since all the participants were coming from the same company and had overall knowledge about technology and IT industry. They were aware of avatar robots and what benefits they could bring in the near future. Some of the participants said they already used avatar in different business meetings during Covid-19 pandemic and had excellent feedback from their customers and business partners. From the calculation performed using RStudio, the level of experience participants already had with avatar robots: $M =$ 3.4 and $Mdn =$ 4.0 *(1 = no experience; 7 = expert user in this field).*

A few participants with some experience on avatar robots used a Double 3 telepresence robot [17] developed by Double Robotics previously. This was different from CommU since it was a physical robotics machine and offered another set of features. CommU digital robot used for this research was only in virtual format and participants noticed a few benefits as digital robots. In the first place, the virtual solution was free and easy to install in comparison to Double 3 telepresence robot which costs 6000$ [17]. Another reported benefit of virtual solutions is reliability. Participants reported that if errors on physical avatar are found they would need to of course contact the support team to find the issue and even sometimes they would need to send the whole telepresence avatar robot to manufacturer to get repaired. Situations like these could take up to a few months until a solution is found and presented huge issue. On the other hand, software, and virtual solutions like virtual CommU could be fixed in a much shorter period if any issues could occur.

4.2 Satisfaction During the Virtual CommU Conference Session

The overall satisfaction reported by the participants during the virtual meeting session while using CommU Conference Platform: $M = 4.556$ and $Mdn = 5.0$ *(1 = very bad; 7 extremely good).* Generally, most of the participants reported they were satisfied with the overall meeting experience while using CommU Conference platform. During one meeting participants experienced an awfully bad echo effect and had to switch from muted to unmuted state while trying to speak one to another. Despite there already was a well-accepted rule of one muting while s/he is not speaking [18], it caused only small problems in conversation. The echo issue seemed to appear only during one meeting session and could also be a result of faulty equipment like headphones and microphones. Simply explained, before participants join the CommU Conference call, they need to be sure their equipment is working without any problems. Participants also reported they appreciated the moving gestures from the CommU virtual avatars although it was almost equivalent benefits to "flashing icon" which indicated very nicely when someone is speaking or every special separate icon for sending a request to speak that other video call platform like Zoom or Microsoft Teams offer.

Furthermore, CommU digital avatars are presented in a baby shape and some participants expressed their unwillingness to use this format for important online consultations, sales or virtual sessions with potential new clients. Despite baby like features of the digital avatars, it was possible to use the system efficiently because the team members already knew each other, and the atmosphere of the call was not strictly formal.

4.3 Frequency of Stress During the Video Meeting Sessions

Previously, during Covid-19, all the participants had experienced working fully from home and high usage of video conferencing tools [2], which they had to adapt to the constant communication and collaboration via Zoom, Slack, Microsoft Teams, and similar applications. During the interview discussions with the participants some of them expressed how stressful it can be to turn on the camera for the video call and that sometimes it was not so easy to find a perfect spot at home for video conferencing. Especially for those participants with small children, it was difficult to explain to their children that they should not enter the room during important online video meetings.

Concerning about the stress level, the survey result shows: $M = 3.333$ and $Mdn = 3.0$ *(1 = not stresses at all; 7 very stressed).* This result indicates that participants felt some levels of stress while participating in the normal video calls such as Microsoft Teams.

Similar result can be also assumed for other well-known video conferencing platforms like Zoom or Slack. It is hard to make direct comparisons of stress levels experience in well-known video conferencing platform to CommU Conferencing application, however, the following results show that there is a potential for CommU Conferencing application to help participants with anxiety issues to feel less stressed during important online meeting sessions. In our sessions, two participants answered that they directly felt how using CommU Conferencing platform helped them feel lower levels of stress during virtual meeting session, and majority, six participants reported mixed feelings but felt some small different in improvements of their levels of stress. While one participant reported he did not feel any difference in levels of stress, those could easily be those who usually are not experiencing any stressful situations during the video meeting sessions.

The participants also reported to see further potentials in using CommU Conference application. When they don't know other participants, they have a challenging time introducing yourself to a new group of people. Similarly when the situation is stressful, like an important meeting session with senior management about high-priority clients or a deal that needs to be closed, they felt challenged. The seven participants have reported that they would prefer to use CommU Conference application with people they haven't met before, while the rest two participants who would prefer to use CommU with already known participants.

4.4 Perception of Social Presence While Using CommU Conference

The participants reported both in the interview and the survey that using CommU conference application made some difference in perception of the social presence of other participants who joined the meeting session. In the analysis of the responses gathered from the survey, the results were as follows: $M = 3.889$ and $Mdn = 4.0$ *(1 = no difference at all; 7 huge difference).*

The results indicated there was not a drastic change, but overall, some improvements regarding the perception of social presence for the participants. This result was supported by the interview. A few participants mentioned they appreciated gestures produced by the digital avatars, which drastically improved communication in online environment. Another participant commented that the head movements of the digital avatar and avatars' ability to focus the attention on the participant who is speaking at that exact moment proved extremely beneficial to all the participants. Hand movement of the digital avatar was also reported beneficial as compatible with "flashing icon" functionality which is available in well-known video conferencing applications.

4.5 Trust Establishment While Using CommU Conference

The analysis of the survey results shows that participants performed an exceptionally low impact on the trust establishment. This is not surprising as trust establishment is a very complex process and much longer and frequent sessions should have been organized to get meaningful results. Although trust establishment could be a very important aspect for further research in similar areas, it was not a principal area for this research. Results were as follows: $M = 2.222$ and $Mdn = 2.0$ *(1 = no difference at all; 7 huge difference).*

5 Discussion

The main purpose for this research was to investigate whether new conference systems like CommU platform and similar digital avatar robotic solutions can make any significant impact on the perception of social presence and co-presence for participants and bring overall improvements during video meetings in the virtual environments. The pandemic showed how the global situation could easily shift and completely change the working environment for millions of workers in many different industries, thus, the importance of remote work and virtual meeting session should not be neglected.

The data presented in the previous section indicate the participants saw a potential in using such technological platforms and can see several benefits. Primarily, our analysis indicates that CommU Conferencing platform would be beneficial in virtual meetings for those people who are experiencing higher levels of stress and have harder times in presenting important topics to people positioned in the upper management hierarchy and people they have not met previously. There would be a couple of reasons why the participants considered improved perceived social presence and co-presence.

First, the participants achieved improved perceived social presence and co-presence as all participants could situate at the same stage as equal appearance as digital avatar. Different from the typical conference system, the participants at CommU could recognize themselves visually as a part of the meeting participants in the virtual CommU meeting room, where all participants present and visible each other. While digital conference sessions with ordinary online conference systems such as Zoom and Teams, sometimes participants neglect each other by accident or forget some participants as the system hid away from the screen. It was not within our research scope, but the context of hybrid meetings would be more complexed as participants in satellite location can easily fall away from the discussion among the co-located participants.

Second, our results indicate that digital video conferencing systems could be especially beneficial to young age group and unexperienced workers during job interviews or for important meeting sessions with senior management. This was also evident especially comments from interviews such as “feeling safe under the condition, which requires no need to show faces”. Younger workers and the novice workers could have important benefits during virtual meeting sessions since they would not be perceived by their looks as young and unexperienced. Those biased opinions could sometimes be formed by senior managers and more experienced workers in the company. It can be easily imagined that by not forming those biased opinions during the initial video calls and introduction sessions, more important tasks and challenges could be given also to a younger employee. Age could be just one aspect of varied biased opinions, and this could be applied to many other areas in the work settings such as: gender, race, nationality, and others. Additionally, by using CommU, digital avatar for video meeting sessions, the participants are all perceived the same and have the same physical shape in the virtual environment. In this situation, all participants in meeting sessions, will be primary focused on the matter of the issue and on the main topic of the conversation. This can result in overall better and more efficient meeting experience and at the same time limit the formation of biased opinions which could create an unpleasant meeting atmosphere.

Participants did not experience any major difference in trust establishment, but they expressed that social presence was perceived better while comparing it to well-known video conferencing platforms. There was no difference in the age groups about the acceptance of such technologies, and all the participants were aware of the potential benefits we could have in the future by using such virtual teleconferencing systems.

6 Conclusion

In this study, in order to understand effects of digital avatar on perceived social presence and co-presence in business meetings between the managers and their co-workers, we conducted a preliminary experiment. Our analysis showed that participants experienced improvements in perception of social presence and co-presence in virtual meetings. Our results suggest that participants appreciated having digital avatars and saw benefits of implementing human gestures to the digital avatar for improved communication in distance. It is important to mention that embodiment of the digital avatar might matter to the use contexts, such as business or casual meetings. The baby shape format was a nice option to have for casual and relaxed meetings, such as weekly meetings between manager and worker like our case.

The participants are fond of using the digital avatars and did not experience any negative sides while having one to one meeting sessions. Most of the participants reported that they felt some improvements in perception of social presence than their usual virtual meetings. Minor technical problems occurred for some participants but those were resolved by changing the headset or the microphone and in general all participants were positively impressed with CommU Conference platform. All in all, participants felt such conferencing solutions could easily become a primary way of virtual communication in the future and could forecast virtual meeting experiences could be drastically improved by using digital avatar systems. One of the most unexpected indications of this research is that such digital avatar conferencing systems can reduce formation of biased opinions towards other participants that we have never met and encountered before. By using avatar in meeting contexts, all the participants can primarily focus on the content of the meeting session and have a better and more fair understanding of the issues discussed during the meeting session.

6.1 Limitations and Future Work

There are a couple of limitations in this research. The biggest limitation is the size of the participants and the duration of the experiment. The participants are small with five pairs in total, and the field experiment period was short with a few sessions in three months. Originally, the research was planned not as lab experiments but as real-world field experiments at the real work environment as a preliminary investigation. It is a great advantage to conduct experiments under non-fictious setting as we can investigate genuine impacts of digital avatar on remote work by running longitudinal experiments. This article reported only the preliminary part of the experiment, however, we are currently planning to conduct more meeting sessions and follow-up meeting sessions with the participants to observe changes of participants' mindset.

Another potential limitations of this research could be the location where the research was carried out. Denmark, together with other Nordics region in general are overall one of the most developed and digitalized regions in the world [19]. All the participants of our study easily accepted the usage of such an innovative technology and were really interested in learning more about it. This is an incredibly unique situation and could potentially result in accumulating optimistic data and should be interpreted with caution. In other regions, which are not as technically developed, it would be interesting to see the effects of such technologies and whether the workforce would accept new technology such as virtual avatar.

The same caution should be paid regarding the industry. The workers in the IT industry, in which the research conducted, understand the technology and its potential needs to get their job done. Some participants had experience with avatar robots. Further research should be conducted with participants coming from other industries and from countries with a lower level of digitalization.

There are some potentials of future works. From observations gathered during the virtual meeting sessions with participants, it was evident that participants were not fully immersed in the virtual experience. They were not immediately aware of the benefits and differences CommU Conferencing platform can bring and the primary advantages of using such platforms. Thus, it would be highly suggested to further improve and advance CommU Conferencing platform to use virtual reality as its main interface. Using CommU Conferencing platform with web browser on a laptop or desktop computer was functional and participants got overall understanding of the benefits. However, huge improvements could be generated if the entire system was available on a VR platform. By implementing such a solution on a VR platform, participants would be fully immersed in the virtual meeting environment, and it would be much easier to understand the benefits and potential that can be unlocked. Something similar can be seen by the development of Metaverse – a virtual reality environment where in the future people will be able to work, socialize and in general do everything we are doing currently in our real lives. Using technological development, such technologies will be possible soon if we, human beings, accept such technologies and a way of living. Today, we are not ready and there are a lot of thought given on whether such technologies are needed and what kind of benefits they could bring, but concepts like Metaverse are increasingly being mentioned and a lot of giant IT corporations are starting to develop solutions to support such technologies in the future.

Digital avatar robots like CommU and other similar avatar robotic solutions will be used increasingly as the time goes by and as the people get used to the existence and potential benefits of such robotic systems. Currently, most people coming from technology industries are perceived to have some benefits from using such robotics systems because of flexibility in their work. They are also the workers who understand its crucial role in transforming the global workforce and the way to work in the future. Todays' workforce is heavily influenced and dependent on global economic, political, and financial decisions and without effective usage of technology, it would be extremely hard to stay productive and deliver better results. Because of that, every industry and every job should be aware of the benefits that can be achieved by understanding technology and whose technological advancements can help run a better business.

The digital technology mentioned in this research is still being developed and still very new to a substantial number of people. At the same time, it could potentially be something to make drastic changes in overall meeting experience and also in the perceived perception of social presence and co-presence of other participants. Still, we are witnessed how such avatar robotic systems are being also used in other industries and their application constantly grows. As seen from the effects of the Covid-19 pandemic, we are very depended on the technology and technological advancements and today's modern workforce would not be able to be as productive as it currently is if the benefits of modern technologies are not being used to its full potential. From the analysis of the results, it was evident how such digital avatar robotic solutions could improve perception of social presence, co-presence and improve overall meeting experiences. However, still a lot of research should be done with this technology to completely understand its benefits in varied situations and environments.

Acknowledgements. We would like to thank reviewers for taking an effort to review the manuscript. This work was supported by JST Moonshot R&D Grant Number JPMJMS2011.

References

1. Spiteri, G., et al.: First cases of coronavirus disease 2019 (COVID-19) in the WHO European region, 24 January to 21 February 2020. Euro Surveill. **25**, 9 (2020). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7068164/
2. Molla, R.: The pandemic was great for Zoom. What happens when there's a vaccine? Vox (2020). https://www.vox.com/recode/21726260/zoom-microsoft-teams-video-conferencing-post-pandemic-coronavirus
3. Yang, L.: The effects of remote work on collaboration among information workers. Nat. Hum. Behav. **6**(1), 43–54 (2021). https://www.microsoft.com/en-us/research/publication/the-effects-of-remote-work-on-collaboration-among-information-workers/
4. Maison, D., Jaworska, D., Adamczyk, D., Affeltowicz, D.: The challenges arising from the COVID-19 pandemic and the way people deal with them. A qualitative longitudinal study. PLoS ONE **16**(10), e0258133 (2021)
5. Gleason, B., Greenhow, C.: Hybrid learning in higher education: the potential of teaching and learning with robot-mediated communication. Online Learn. J. **21**(4), 159–176 (2017)
6. Top 30 Information Technology companies in Europe by sales in 2018 (2018). https://www.globaldatabase.com/top-30-information-technology-companies-in-europe-by-sales-in-2018
7. SAP: Health & well-being: a social and cultural perspective. https://www.sap.com/about/company/purpose-and-sustainability/social-responsibility/mental-health.html
8. Bainbridge, W.A., Hart, J.W., Kim, E.S., Scassellati, B.: The benefits of interactions with physically present robots over video-displayed agents. Int. J. Soc. Robot. **3**, 41–52 (2010). https://doi.org/10.1007/s12369-010-0082-7
9. Bente, G., Ruggenberg, S., Kramer, N.C.: Social presence and interpersonal trust in avatar-based, collaborative net-communications. In: PRESENCE 2004: Proceedings of the 7th Annual International Workshop on Presence, pp. 54–61 (2004)
10. Jung, Y., Lee, K.M.: Effect of physical embodiment on social presence of social robots (2004)
11. Lee, K.: Presence, explicated. Commun. Theory **14**(1), 27–50 (2004)
12. Mende, M.A., Fischer, M.H., Kühne, K.: The use of social robots and the uncanny valley phenomenon. In: Zhou, Y., Fischer, M.H. (eds.) AI Love You, pp. 41–73. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-19734-6_3

13. Brown, G., Prilla, M.: The effects of consultant avatar size and dynamics on customer trust in online consultations. In: Proceedings of the Conference on Mensch and Computer, pp. 239–249 (2020)
14. Haring, K.S., et al.: Robot authority in human-robot teaming: effect of human-likeness and physical embodiment on compliance. Front. Psychol. **12**, 625713 (2021)
15. Fong, T., Nourbakhsh, I., Dautenhahn, K.: A survey of socially interactive robots. Rob. Auton. Syst. **42**(3–4), 143–166 (2003)
16. Rae, I., Takayama, L., Mutlu, B.: The influence of height in robot-mediated communication. In: 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 1–8. IEEE (2013)
17. Double Robotics – Telepresence Robot for the Hybrid Office. https://www.doublerobotics.com/double3.html
18. Zoom Call etiquette – Academic Technology Call Center. https://athelp.sfsu.edu/hc/en-us/articles/360044451294-Zoom-call-etiquette
19. Copenhagen Capacity: Denmark is the most digital country in the world. https://www.copcap.com/news/denmark-is-the-most-digital-country-in-the-world (2018)
20. Tanaka, K., Nakanishi, H., Ishiguro, H.: Comparing video, avatar, and robot mediated communication: pros and cons of embodiment. In: Yuizono, T., Zurita, G., Baloian, N., Inoue, T., Ogata, H. (eds.) CollabTech 2014. CCIS, vol. 460, pp. 96–110. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44651-5_9

Presentation Method for Conveying Nonverbal Information in Online Conference Presentations with a Virtual Stage

Hiroki Echigo[1](✉), Kanan Abe[1], Yuki Igarashi[2], and Minoru Kobayashi[1]

[1] Meiji University, 4-21-1 Nakano, Nakano-Ku, Tokyo, Japan
{hiroki.echigo,kanan.abe}@koblab.org, minoru@acm.org
[2] Ochanomizu University, 2-1-1 Otsuka, Bunkyo-Ku, Tokyo, Japan
yukim@acm.org

Abstract. The recent spread of coronavirus (COVID-19) has meant that online conference presentations are becoming more and more frequent at national and international level. We believe that these online presentations will remain an option even after the pandemic has subsided. One of the challenges of online conference presentations is that it is difficult to convey nonverbal information such as gestures and the facial expressions of the presenter. In this paper, we propose the "Stage-like Presentation Method", which involves projecting the whole body, and investigates how the presence or absence of nonverbal information from the presenter affects the audience. A comparison of the proposed method with two other presentation methods confirmed that the audience considered it the most effective. The method was used by seven people in actual conference presentations, and it was found that the audience's impressions changed according to the details of the setting. This research confirmed that the Stage-like Presentation Method left the audience at online conferences with a good impression of presentations. It also suggests that audiences find visual nonverbal information useful.

Keywords: Presentation · Nonverbal communication · Online conference

1 Introduction

The number of online conference presentations has increased rapidly around the world in recent years. This is due in particular to the recent spread of coronavirus (COVID-19). As a result, CollabTech2020 and CollabTech2021 were held online only. Online conference presentations have the advantage that participants can be anywhere in the world, and do not need to be physically present at the conference venue. Even after the pandemic has subsided, online conferences are predicted to continue for presenters who are unable to be physically present for any reason. Examples of potential beneficiaries include people with physical disabilities, pregnant women, people caring for young children, those with childcare issues and people involved in long-term care. Online conferences may also be held in future as a response to localized issues such as earthquakes, wars and natural disasters.

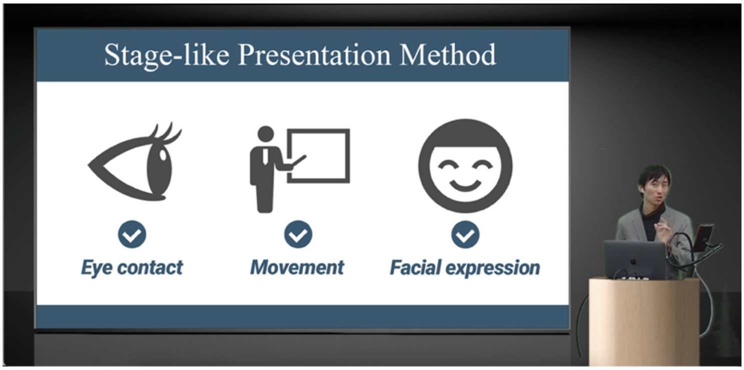
L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 98–111, 2022.
https://doi.org/10.1007/978-3-031-20218-6_7

Conference presentations are an important opportunity to convey research content to the audience. Presentation methodology is described in a number of works [1, 2] which suggest that, in addition to the content, structure and design of the presentation, the quality can be enhanced by clues from nonverbal communication such as the presenter's appearance, gestures, eye movements and facial expressions. However, online conference platforms such as Zoom [3], Webex [4], Microsoft Teams [5] and Google Meet [6] often display the presentation slides in the largest part of the screen and the face of the presenter in a small window. This makes it hard to transmit nonverbal information properly to the audience.

To resolve this problem, in this paper, we propose the "Stage-like Presentation Method" (SPM), which gives the impression that the presenter is making the presentation on a stage (see Fig. 1). The SPM can convey nonverbal information to the audience, such as the presenter's eye movements, facial expressions and gestures.

We investigated how the presenter's nonverbal information affected the audience in online conference presentations. We first prepared the three-presentation method shown in Fig. 2, and conducted experiments to investigate the impression the presentation made on the audience. The experiment showed that the audience considered SPM the most effective of three types of presentation method. We also obtained results suggesting that visual nonverbal information, such as the presenter's eye movements, whole-body movements and gestures, were factors which left the audience with a good impression.

The SPM has been used by seven students in actual conference presentations so far. It was found in using the SPM that the impression of the audience was affected by the presenter's eye movements and detailed settings. Therefore, we focused on three points: "the presenter's eyes", "body orientation" and "balance between the presenter and the podium", and conducted an additional survey. The investigation demonstrated that the body orientation of the presenter should be the same as the orientation of the slide, and the presenter's eyes should preferably be facing the front of the screen.



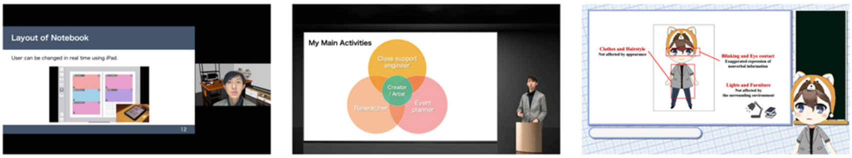
Stage-like Presentation Method

Eye contact

Movement

Facial expression

Fig. 1. Stage-like presentation method.



(a) Web Conferencing Method (b) Stage-like Presentation Method (c) VTuber Presentation Method

Fig. 2. Presentation methods.

2 Presentation

Mehrabian [7] reports that "visual information such as appearance" influences 55% of human judgment, "auditory information such as voice and speaking style" influences 38% and "linguistic information such as spoken language and content" 7%. Therefore, a number of works [1, 2] which describe presentation methods include not only the content of the presentation, but also considerable nonverbal information, such as the tone of the voice, the clothes worn, facial expressions, eye movements and gestures. The presentation methods described in these works involve knowledge and a number of techniques that can be applied to online conference presentations. We believe there is virtually no difference between face-to-face conference presentations, and auditory and linguistic information such as speaking style, tone of voice, slide composition and easy-to-read slide creation. In terms of visual information, however, since listeners cannot see the presenter, it is difficult for the presenter to make eye contact with listeners, or to move their body in response to listeners' eye movements and gestures. Indicating enthusiasm by incorporating gestures can also be difficult. In addition, the reaction of the audience is unknown, and it is difficult for the presenter to understand whether the content of the presentation is being conveyed adequately, and whether it is making a good impression.

Presentation sensei [8] can give feedback to the presenter in real time, for example concerning the speed of speech, tone of voice, and face orientation. Chen et al. [9] are analyzed speaking skills with a multimodal corpus. Ishino et al. [10] have proposed a lecture robot, which reconstructs lecturer's presentation behavior in a lecture, showing that gestures and paralanguage are important. As stated in these studies, good presentations involve multiple factors. However, few studies have investigated the audience's impression of visual nonverbal information.

In this paper, we focus on nonverbal visual information in online presentations, and investigates the extent to which it affects the audience.

3 The Stage-Like Presentation Method

Figure 3 shows the SPM. The SPM is produced using OBS Studio [11], software which enables live distribution and video recording. This section will provide separate accounts of preparing a presentation for a real-world situation and preparing a presentation in OBS Studio.

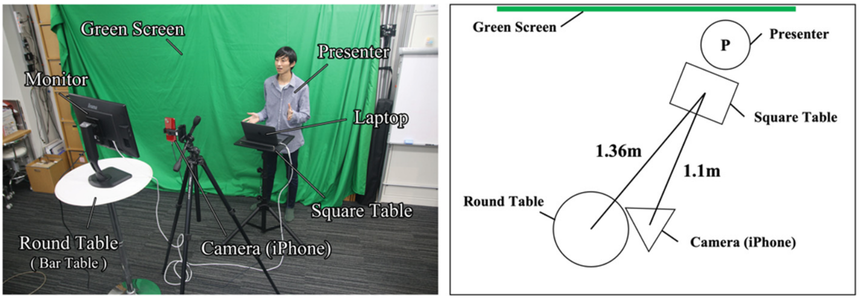


Fig. 3. Using the stage-like presentation method (in a real-world situation).

3.1 Preparing a Presentation for a Real-World Situation

A green screen is first erected behind the presenter (see Fig. 3). A square table with a laptop is placed in front of the place where the presenter will stand. A camera and a monitor are then placed in front of the presenter's line of sight, and each is connected to the presenter's laptop. The presenter can see the slides during the presentation by viewing them on the monitor.

3.2 Preparing a Presentation in OBS Studio

In OBS Studio, multiple images, as well as the presenter's video transmitted from the camera, are superimposed to make the screen as shown in Fig. 1. The chroma key makes the video transparent against the green background. The stage and podium on the screen are produced by Cinema 4D [12], 3D CG software, and are available in Unity [13] (Fig. 4). If the 3D model is used as it is, processing will be heavy and the composition of the camera image will be complicated. Therefore, the 3D model viewed from an arbitrary position is produced as an image and displayed in OBS Studio. The image of the stage and the image of the podium are separated, and the layer is set so that the image from the camera comes below the image of the podium (Fig. 5). These images used in this method are publicly available for free by the author [14]. The presence of the podium image gives a sense of unity to the entire screen.

4 Comparative Experiment

This section explores three presentation methods, the SPM shown in Sect. 3, the conventional Web Conferencing Method (WCM) and a presentation method using avatars (VTuber Presentation Method (VPM)). Comparative experiments were conducted to see how visual nonverbal information affected the audience. In all presentation methods, the presentation was recorded using a laptop (Apple MacBook Pro 13 inch, Apple M1 chip built-in type), and a sub monitor was used for screen sharing (Dell SE2416H). The slides used for the presentation were made with PowerPoint [15].

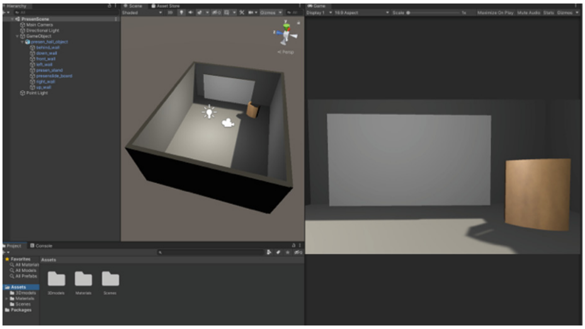


Fig. 4. 3D model of the stage in unity.

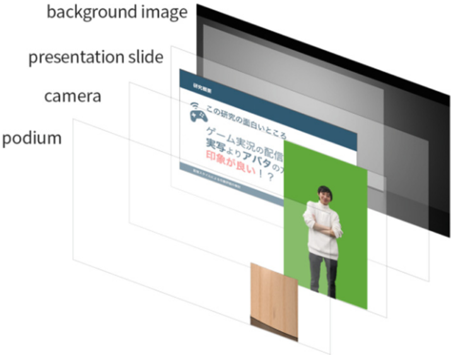


Fig. 5. Configuration in OBS Studio.

4.1 Preparing the Experiment

The Web Conferencing Method (WCM). Figure 6 shows the presentation method of the WCM. The WCM represents the method commonly used on Web conferencing systems, such as Zoom [3], Webex [4], Microsoft Teams [5] and Google Meet [6]. The presenter uses the built-in camera on the laptop to display his / her face and makes a presentation into this camera. It displays the presenter and the slides side-by-side, and we did not use the functions superimposing the presenter in front of the slides, which some of those Web conference systems have. The range of the presenter reflected in the camera is currently limited, so even if the presenter makes gestures as he or she speaks, as shown in Fig. 6, only the face above the shoulders is transmitted to the listener.

The VTuber Presentation Method (VPM). Figure 7 shows the presentation method of the VPM. A VTuber is a person who distributes on a video distribution site using a virtual avatar drawn in 2D CG or 3D CG. VTubers originated in Japan, and the practice is gaining rapidly in international popularity. Research on VTuber is also being conducted internationally [16]. The VPM only transmits nonverbal information such as the facial movements and expressions, eye movements and blinking of the person operating the avatar. Information such as the operator's beard or the surrounding environment,

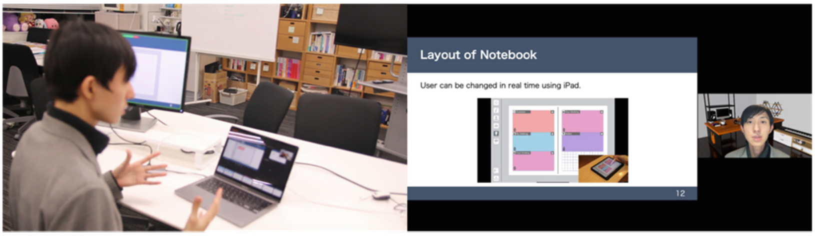


Fig. 6. Web conferencing method.

for example, is not transmitted to the listener. Therefore, listeners can see nonverbal information easily, and it is considered to be an excellent method for transmitting this type of information. In recent years, there has been research on communication using avatars as well as promotion of products which use them [17, 18]. This prompted us to compare these with the SPM.

The presenter's 2D CG avatar creates an original picture with CLIP STUDIO PAINT [19], which is illustration-production software, and creates a motion model with the data in Live 2D [20]. By importing this model into VTube Studio, facial expressions and body movements acquired by the iPhone's front camera can be applied to 2D CG avatars in real time. The 2D CG avatar video is transferred from the iPhone to the laptop in real time. The settings of OBS Studio are almost the same as the SPM described in Sect. 3.

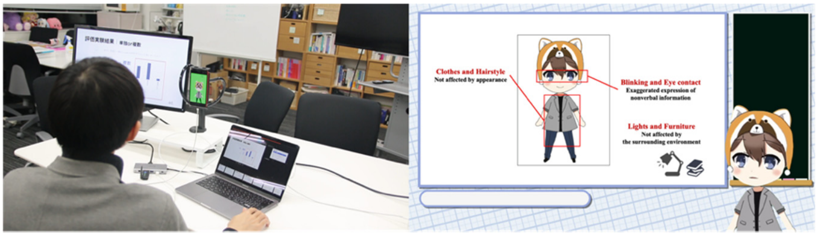

Clothes and Hairstyle
Not affected by appearance
Blinking and Eye contact
Exaggerated expression of nonverbal information
Lights and Furniture
Not affected by the surrounding environment

Fig. 7. VTuber presentation method.

4.2 Experimental Environment and Conditions

We prepared presentation videos of about five minutes for each of the three presentation methods: the WCM, the SPM and the VPM. Participants in the experiment were asked to watch the three presentation videos one by one. The presentation videos involved participants' own content from past academic conferences, but each was condensed to about five minutes. Table 1 shows the questions which were put to the participants in the experiment. As they watched each video, they answered the questions.

Q1 and Q2 in Table 1 were evaluated on a 5-point Likert scale. The same presenter was in charge of all three conditions. The aim was to minimize the difference in impressions resulting from anything other than visual nonverbal information, such as voice or speaking style. The Latin square method was used to consider the effect of the order on the result.

After the experiment, the participants were asked the questions shown in Table 2. There were 20 participants aged 19–26 (fourteen males and six females). They were not informed in advance about the presentation method represented by each of the three conditions.

Table 1. Questionnaire used during the experiment.

No.	Question
Q1	i) Did you feel that the presenter looks at the audience?
	ii) Did you feel the presenter's movement?
	iii) Did you feel the presenter's facial expression?
	(1. Not at all. – 5. Very much.)
Q2	How did you feel about the presenter?
	(1. Not very good. – 5. Very good.)
Q3	Please enter the reason for your response to Q2

Table 2. Post-experiment questionnaire.

No.	Question
After-Q1	Which of the three presentation methods gave you the best impression of the presenter?
	(The WCM, The SPM, The VPM)
After-Q2	Please enter the reason for your response to After-Q1

4.3 Results

Presenter's Eye Contact, Movement and Facial Expression. Figure 8 (a) shows the results of question Q1-i) "Did you feel that the presenter looks at the audience?". For all three methods, evaluation values were selected from 1 to 5, and although there was almost no difference between the WCM and the VPM, the SPM had a higher average evaluation value than the other conditions. A Wilcoxon signed rank test confirmed a significant difference between the WCM and the SPM at the 1% level, and a significant difference between the SPM and the VPM at the 5% level.

Figure 8 (b) shows the results of question Q1-ii) "Did you feel the presenter's movement?". Many of the participants in the experiment gave the SPM an evaluation value of 5, higher than for the other methods. When a Wilcoxon signed rank test was performed, a significant difference was confirmed at the 1% level between the WCM and the SPM, the WCM and the VPM, and the SPM and the VPM.

Figure 8 (c) shows the results of question Q1-iii) "Did you feel the presenter's facial expression?". Evaluation values were selected from 1 to 5 for all three methods, and there was almost no difference between the WCM and the VPM. The SPM had a higher average evaluation value than the other methods. A Wilcoxon signed rank test confirmed a significant difference at the 1% level between the WCM and the SPM, and a significant difference at the 5% level between the SPM and the VPM.

Though the WCM had a larger facial image than the SPM, the participants reported that it did not give a good impression to the participants because the screen is split into multiple windows of presentation slides and camera images.

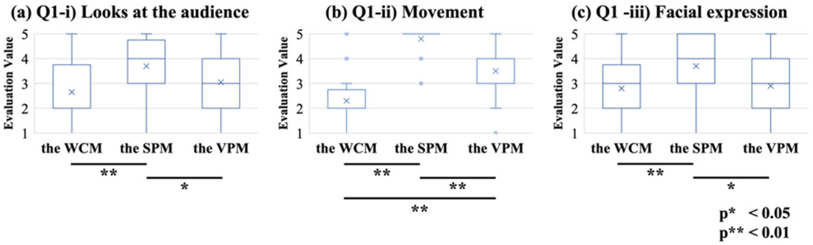


Fig. 8. Results of questionnaire Q1.

Impressions of the Presenter. Figure 9 shows the results of question Q2: "How did you feel about the presenter?". Many of the participants gave an evaluation value of 3 for the WCM. Evaluation values for the SPM and the VPM were between 3 and 5, and the average evaluation value was high. A Wilcoxon signed rank test showed a significant difference between the WCM and the SPM, and between the WCM and the VPM at the 1% level. The participants who gave the WCM a rating of 3 said, "I felt I was speaking in a monotonous way" or "I thought it was a common presentation method in online conferences". Many of the participants who gave the SPM a value of 5 said that "the content was easily transmitted by presenter's hand movements and eye gaze" and "considerable heat was transmitted from the presenter's eyes and gestures". On the other hand, the participants who gave a value of 3 said, "I could feel the movement well, but I couldn't focus on the presentation slide because I was paying too much attention to the movement of the presenter." Many of the participants gave the VPM an evaluation value of 4. There were a number of opinions about 2D CG avatars. These included: "As the presenter was an avatar, I was able to hear the presentation in a charming and moving atmosphere", and "As the presenter was an avatar, I was left with a pleasant feeling". On the other hand, the participants who gave an evaluation value of 3 said, "It was harder to recognize changes in movements and facial expressions than it would have been in real humans, and it was harder to feel emotions".

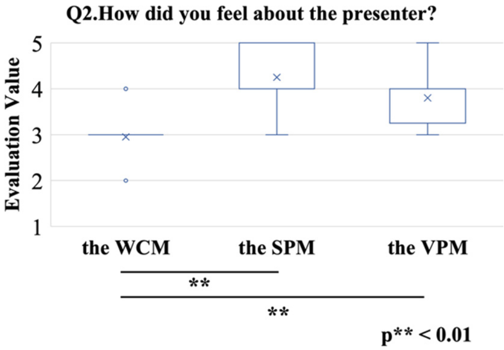


Fig. 9. Results of questionnaire Q2.

Best Presentation Method. Figure 10 shows the results of the question "After-Q1", asked after the experiment: "Which of the three presentation methods gave you the best impression of the presenter?". The number of participants who answered the SPM was the highest at 70%, followed by 30% who preferred the VPM. None of the participants chose "no difference" in terms of the WCM. The participants who chose the SPM said, for example, "Since the gestures are similar to the presentations at the actual physical venue, there was a sense of presence", or "the movement of the presenter meant I didn't find the presentation so monotonous, and I could listen to it without getting tired". One of the participants who chose the VPM said, "Since both slides and the avatar are two-dimensional information, I could easily capture all the information about the presentation and the presenter at the same time, just like looking at a single picture".

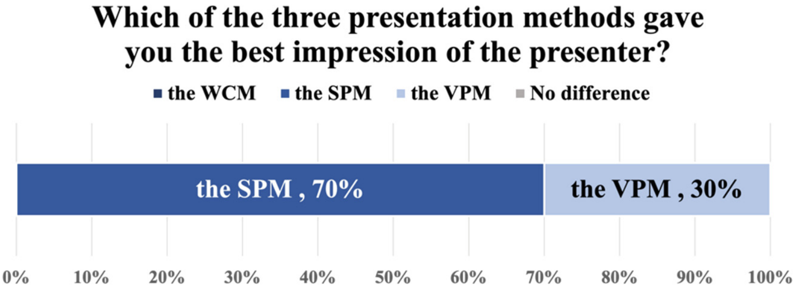


Fig. 10. Results of the questionnaire given after the experiment.

5 Use in Actual Online Conferences

The SPM was used at an actual conference presentation by seven students. Figure 11 shows the state of the presentation. One of the students made three presentations at separate online conferences using the proposed method, and received two presentation awards. The participants at one of the conferences used the Slack communication tool, and their impressions included: “A wonderful presenter screen”, “The method is effective because of the clear gestures” and “What a nice presentation method!”.

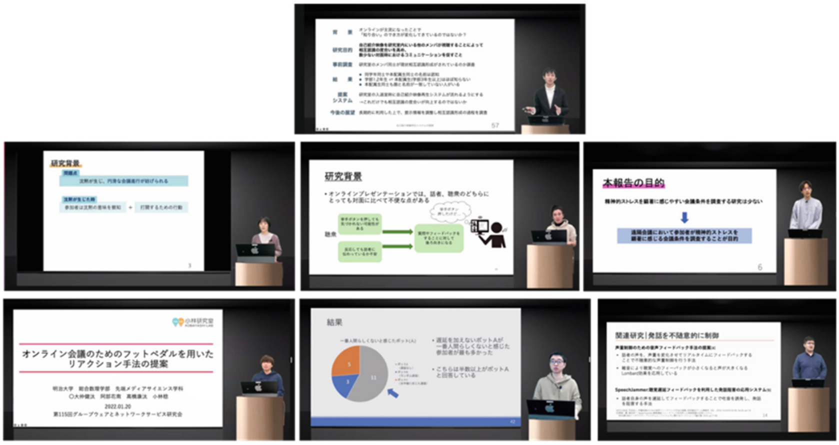

研究背景
研究背景
本報告の目的
オンライン会議のためのフットペダルを用いたリアクション手法の提案
結果

Fig. 11. Screenshots of presentations by seven students using the proposed method at an online conference.

6 Additional Research

The use of the SPM by these seven people highlighted a number of issues with the method. The results detailed in Sect. 4 show that the SPM makes it easy to convey the presenter’s eye and body movements, so that listeners are left with a good impression. Nonverbal information is well communicated, and even the smallest adjustments can change the impression of the audience. Comments included the following: “Would it be better for the body to face the slide?”, “Would it be better for the presenter to look at the camera?” and “What is the proper balance and composition between the presenter and the podium?”. In fact, people sometimes felt uncomfortable when they saw the presenter’s unnatural-looking eyes. We therefore conducted further research focusing on the three issues of “body orientation”, “presenter’s eyes” and “balance between presenter and the podium”.

For this additional research, we prepared four sets of two different images as shown in Fig. 12. Results and discussions are presented for each of the patterns compared. Twenty-six people aged 20–59 (Twenty men and six women) participated in this research.

6.1 Body Orientation

Figure 13 shows the results of this research comparing A: “The body facing the right-hand side of the screen” and B: “The body facing the left-hand side of the screen”. 80.8% of participants chose B. A number of reasons were given, such as that it was “because

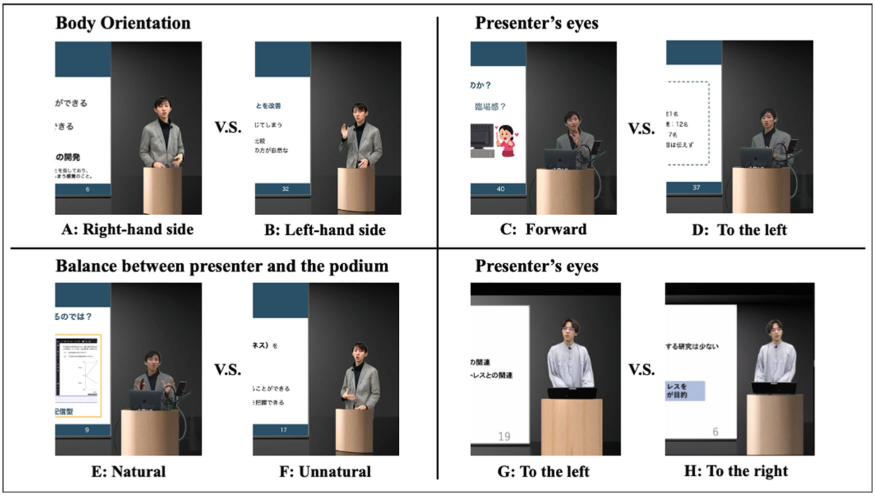


Fig. 12. Image set used in additional research.

the presenter's body is facing the slide side" and "I felt that the presenter's body was facing the slide and the audience". On the other hand, the participants who answered A said, "I felt that the presenter's face was facing me". In other words, it seems that some people care more about the orientation of the presenter's face than the body orientation.

The results indicated that the audience had a better impression when the body was facing the slide side.

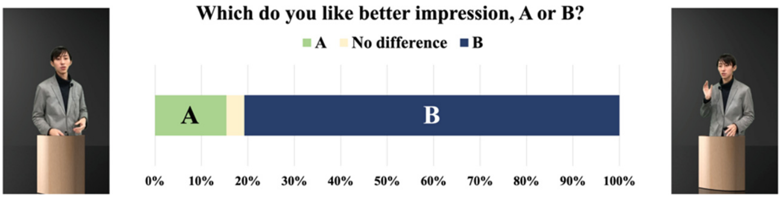


Fig. 13. The results of research on the orientation of the presenter's body.

6.2 Presenter's Eyes (Forward or to the Left)

Figure 14 shows the results of the research comparing C: "The presenter's eyes looking forward" and D: "The presenter's eyes looking to the left-hand side of the screen". Image D is a composition often seen in students who actually presented at the conference in Sect. 5. 80.8% of participants chose C. Reasons for this included: "It felt like the presenter was looking at me", "because the presenter's facial expression is bright" or "because the presenter's gestures can be seen". On the other hand, 11.5% of participants answered, "No difference", and gave reasons such as "I didn't really notice the difference".

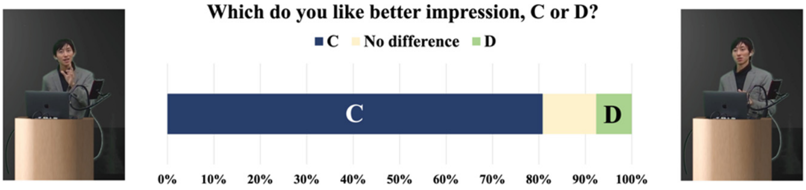


Fig. 14. The result of research on the presenter's eyes (looking forward or to the left).

6.3 Balance Between the Presenter and the Podium

Figure 15 shows the results of research comparing E: "Natural balance" and F: "The podium is small and there is an unnatural balance". 42.3% of participants answered E and 42.3% of participants answered F. The reasons for answering E included: "I felt that the presenter was presenting using slides", "I felt uncomfortable because F had nothing on the podium" and "because the gesture was clear". The reasons for answering F included: "because the presenter seemed to be standing" or "I felt that I preferred the presenter to be standing". Participants who answered "No difference" commented that "both gave a good impression" and "I did not feel any difference".

The results show that some people consider the laptop on the podium to be natural, and none of the participants felt uncomfortable about the balance of synthesis. On the other hand, many participants focused on the gestures and standing postures of the presenters.

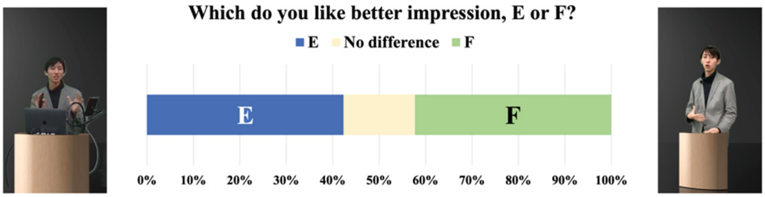


Fig. 15. The result of research on the balance between the presenter and the podium.

6.4 Presenter's Eyes (to the Left or to the Right)

Figure 16 shows the results of research comparing G: "The presenter's eyes facing to the left-hand side of the screen" and H: "The presenter's eyes facing to the right-hand side of the screen". 46.2% of participants answered G, 34.6% answered H and 19.2% answered "No difference". The reasons for answering G included: "I felt like I was in the direction the presenter was looking" and "I felt like the presenter was looking at me". Reasons for answering H included: "I felt that the attitude of the presenter was good" and "I thought the presenter's body orientation was good". Participants who answered "No difference" commented that "I didn't notice the difference" and "both faces were facing forward".

This result indicates that people prefer the presenter's eyes to be looking to the left-hand side rather than the right-hand side of the screen. However, it was also found that many participants focused on the orientation of the body rather than the presenter's eyes.

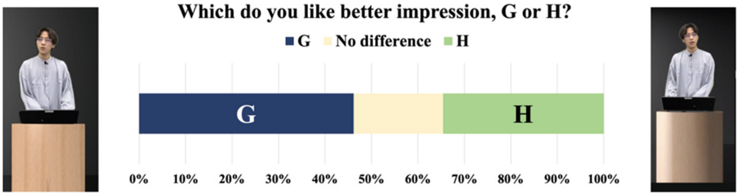


Fig. 16. The result of research on the presenter's eyes (to the left or to the right).

6.5 Discussion of the Additional Research, and Limitations Identified

The results of Sect. 6.1–6.4 indicate that the presenter's body should have the same orientation as the slide, and the presenter's eyes should preferably be looking to the front of the screen. In addition, although this survey was an image, many participants focused on the gestures and facial expressions of the presenter. In this additional research, we therefore wished to investigate "body orientation", "presenter's eyes" and "balance between the presenter and the podium". We therefore prepared the images in this way, so we cannot evaluate the movement of the presenter or the entire presentation.

In the future, we wish to see the SPM used by a wide range of presenters, to evaluate the impression the audience has of the presentation when body movements and gestures are clear.

7 Conclusion

In this paper, we investigated how the presence or absence of nonverbal information from the presenter affected the audience. To do this research, we have proposed the "Stage-like Presentation Method" for making conference presentations. The method projects the presenter from the top to the waist. A comparative experiment confirmed that the proposed method impressed the audience as a presentation style. Of the nonverbal visual information, it was suggested that gestures and body movements particularly impressed the audience. In addition, the preferred body orientation for the presenter was the same as the orientation of the slide, and the presenter's eyes should preferably face the front of the screen.

Acknowledgments. This work was supported by JSPS KAKENHI Grant Number JP18K11410.

References

1. Reynolds, G.: Presentation Zen: Simple Ideas on Presentation and Delivery. 3rd edn. New Riders (2019)
2. Anderson, C.: TED Talks: The Official TED Guide to Public Speaking: Tips and Tricks for Giving Unforgettable Speeches and Presentations. Nicholas Brealey Publishing (2016)
3. Zoom: https://zoom.us/. Accessed 12 June 2022
4. Webex by Cisco: https://www.webex.com/. Accessed 12 June 2022
5. Microsoft Teams: https://www.microsoft.com/en-us/microsoft-teams/group-chat-software. Accessed 12 June 2022
6. Google Meet: https://apps.google.com/intl/en/meet/. Accessed 12 June 2022
7. Mehrabian, A.: Silent Messages: Implicit Communication of Emotions and Attitudes. Wadsworth (1981)
8. Kurihara, K., Goto, M., Ogata, J., Matsusaka, Y., Igarashi, T.: Presentation Sensei: a presentation training system using speech and image processing. In: ICMI 2007: Proceedings of the 9th International Conference on Multimodal Interfaces, pp. 358–365 (2007)
9. Chen, L., Feng, G., Joe, J., Leong, W.C., Kichen, C., Lee, M.C.: Towards automated assessment of public speaking skills using multimodal cues. In: ICMI 2014: Proceedings of the 16th International Conference on Multimodal Interaction, pp. 200–203 (2014)
10. Ishio, T., Goto, M., Kashihara, A.: A robot for reconstructing presentation behavior in lecture. In: HAI 2018: Proceedings of the 6th International Conference on Human Agent Interaction, pp. 67–75 (2018)
11. OBS Studio: https://obsproject.com/. Accessed 12 June 2022
12. Cinema 4D Homepage: https://www.maxon.net/en/cinema-4d. Accessed 12 June 2022
13. Unity: https://unity.com/. Accessed 12 June 2022
14. Image set for use in Stage-like Presenation Method (author's webpage). Accessed 12 June 2022
15. Microsoft PowerPoint Slide Presentation Software: https://www.microsoft.com/en-us/microsoft-365/powerpoint. Accessed 12 June 2022
16. Lu, Z., Shen, C., Li, J., Shen, H., Wigdor, D.: More kawaii than a real-person live streamer: understanding how the otaku community engages with and perceives virtual YouTubers. In: CHI 2021: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, No. 137, pp. 1–14 (2021)
17. Li, F.L., et al.: AliMe Avatar: multi-modal content production and presentation for live-streaming E-commerce. In: SIGIR 2021: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 2635–2636 (2021)
18. Luo, L., Weng, D., Guo, S., Hao, J., Tu, Z.: Avatar interpreter: improving classroom experiences for deaf and hard-of-hearing people based on augmented reality. In: CHI EA 2022: CHI Conference on Human Factors in Computing Systems Extended Abstracts, No. 318, pp. 1–5 (2022)
19. CLIP STUDIO PAINT: https://www.clipstudio.net/en/. Accessed 12 June 2022
20. Live 2D: https://www.live2d.com/en/. Accessed 12 June 2022

Kano Model-Based Macro and Micro Shift in Feature Perception of Short-Term Online Courses

Daniel Moritz Marutschke[1,2](✉) and Yugo Hayashi[3]

[1] Ritsumeikan University, College of Global Liberal Arts, 2-150 Iwakura-cho, Osaka, Ibaraki 567-8570, Japan
moritz@fc.ritsumei.ac.jp

[2] Xiamen University of Technology, No.600 Ligong Road, Jimei District, Xiamen 361024, Fujian Province, China

[3] Ritsumeikan University, 2-150 Iwakura-cho, Osaka, Ibaraki 567-8570, Japan
y-hayashi@acm.org

Abstract. The perception of e-learning and online courses by students can provide valuable insights into course design and user experience. The online landscape is changing and students experience a growing variety of digital educational materials. One of the experiences are expectations (ex-ante perception) and consumption experience (ex-post perception). They can be evaluated individually as well as in relation to each other. Educational data is typically multivariate and of high dimensionality. The implementation of online courses is often costly, the experimental setup is complex, and management needs technical expertise. For this research, an undergraduate-level online course was set up, which was taken and assessed by students. Using ex-ante and ex-post questionnaire evaluations, a shift in perception of several online course features could be observed. The data was analyzed using the Kano method to measure student satisfaction. Attitudes towards 12 features, including ease of use, multimedia inclusion, account settings, and other specific features were gathered before and after taking the online course. The results of the macro shift (expectations vs. consumption experience) and micro shift (individual student's shift within a requirement) were compared. Findings are discussed and implications for online course preparation and design are presented.

Keywords: E-learning · Kano model · Macro and micro shift · Online course · Student satisfaction

1 Introduction and Previous Research

E-learning and online course design has been a growing business and has also been adopted in higher education.

Many educational questions remain open and where some have been addressed, others remain unanswered or need more in-depth investigations. Previous publications include learning motivation, exploring multiple hypothesis

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 112–125, 2022.
https://doi.org/10.1007/978-3-031-20218-6_8

that tie to student satisfaction [18]. Others investigate the satisfaction to more mechanical implementation, such as medical software usage [19] as a virtual, web-based 3D electroencephalogram to propose design references for future learning platforms. The Kano method to poll students on several e-learning factors was previously used by Dominici and Palumbo [7] to construct a theoretical framework of online course design (ex-ante questionnaire only, without implementation and testing). Another research was published by Wang et al. to investigate a hybrid online/face-to-face course using blended learning [20].

There are other considerations, such as different needs for students depending on their subject matter interest. This is partially addressed in this research by choosing an online course topic suited for the target students. The authors also chose and created the content of the online course that students with minimal required technical knowledge can take and complete the course. This was also done to ensure consistency with future research. Findings are expected to be compared between different fields of study.

How an online course is created and what content is presented to prospective students might impact the research outcome and thus the validity of the results. The significance of subject matter difference has been acknowledged since before e-learning, but has not been the focus of online course design [3,9].

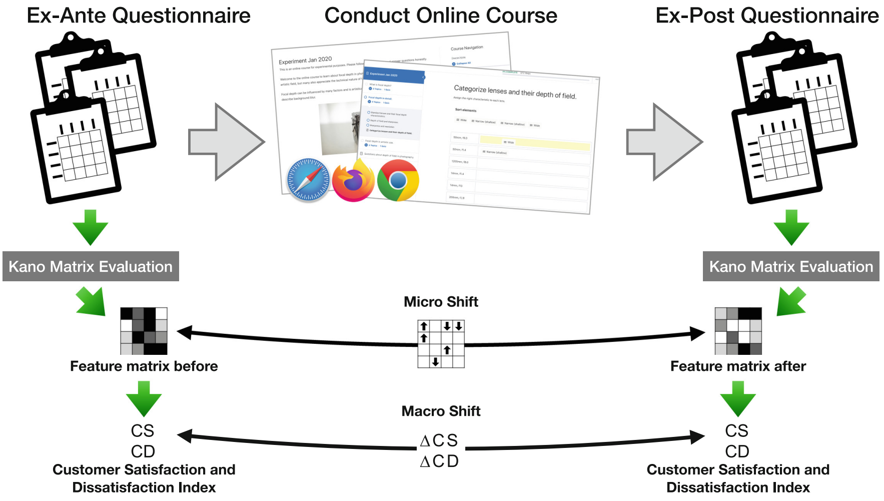


Fig. 1. Experiment overview with questionnaire and online course stages, evaluation using the Kano model, and comparing the macro and micro shift.

This research addresses two perceptual shifts that occur when taking an online course. Questionnaires are conducted and analyzed ex-ante (before) and ex-post (after) the online course. Undergraduate students of the department of comprehensive psychology at the Ritsumeikan university in Japan were surveyed. The experimental setup is shown in Fig. 1. Online course features were gathered via questionnaires and evaluated using the Kano model [8,11]. The ex-ante and ex-post questionnaires are used to determine changes in how features are perceived before taking the online course to after completion. In this research, two shifts are taken into consideration—the *macro shift* indicating the general sentiment of students' satisfaction and dissatisfaction indices; the *micro shift* that investigates individual shifts from Kano model classification (requirement) to another (refer to Sect. 2.1 and Fig. 2). A common approach is to focus on either the theoretical framework or the ex-post effectiveness of e-learning systems, often taking extrinsic motivation into account [6,10,16]. In this paper, the authors propose a framework to address both questions. Results will be analyzed to improve the learning experience of online courses and to investigate possible gaps in expectations. Findings will be assessed to wether the micro shift can inform underlying changes from expectation to consumption experience.

This research addresses the following research questions:

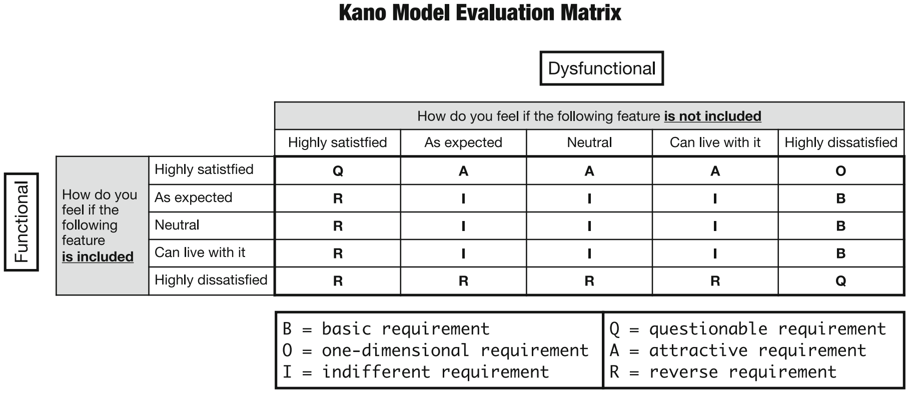
RQ1: Can the Kano model help to identify a shift in expectations versus consumption experience (macro shift)?
RQ2: Can the shifts from RQ1 be categorized using satisfaction and dissatisfaction indices?
RQ3: Can the shift in individual answers (micro shirt) provide additional insight into an ex-ante and ex-post comparison, especially if they cancel each other out in the macro shift?

The structure of the remaining paper has three main parts. Section 2 covers the Kano model (Sect. 2.1) and the approach of this paper to investigate the macro and micro shift in feature perception (Sect. 2.2). Section 3 details the experimental setup (Sect. 3.1), questionnaire design (Sect. 3.2), and the results of this research with the authors discussing implications for online course design using the macro and micro shift (Sect. 3.3). Section 3 also includes lessons learned and limitations in Sect. 3.4 and Sect. 3.5, respectively. The paper concludes in Sect. 4.

2 Kano Method with Perceptual Macro and Micro Shift

The Kano model is a well established tool to conduct customer satisfaction research [8]. At the core is a questionnaire that asks (prospective) users to answer both functional and dysfunctional questions regarding each product's feature. The *functional* question asks how one would feel if a feature is present (implemented) and the *dysfunctional* question asks how one would feel if a feature is missing. Their response is typically rated from highly satisfied to highly dissatisfied for both of the aforementioned functional and dysfunctional question. Combining these answers results in one of six categories of the Kano model.

The sections below describe the Kano model in detail and how the shift from ex-ante to ex-post was analyzed.



Kano Model Evaluation Matrix

Functional		Dysfunctional: How do you feel if the following feature **is not included**				
		Highly satistfied	As expected	Neutral	Can live with it	Highly dissatisfied
How do you feel if the following feature **is included**	Highly satistfied	Q	A	A	A	O
	As expected	R	I	I	I	B
	Neutral	R	I	I	I	B
	Can live with it	R	I	I	I	B
	Highly dissatisfied	R	R	R	R	Q

B = basic requirement	Q = questionable requirement
O = one-dimensional requirement	A = attractive requirement
I = indifferent requirement	R = reverse requirement

Fig. 2. Categories of the Kano model by pairing functional and dysfunctional questions.

2.1 The Kano Model and Terminology

The pair of functional and dysfunctional questions is presented in Fig. 2. Participants have to rate a feature in five levels for both these questions—highly satisfied, as expected, neutral, can live with it, and highly dissatisfied.

The list below details the categories that the Kano model produces. As the original publication describing the categories is written in Japanese, the terminology has been adjusted to reflect this current research. Namely the term *quality* has been replaced by *requirement* to reflect terminology from software engineering and the term *must-be* was replaced by *basic*. The categories are pre-defined by Kano and described individually in the list below.

Most researches that incorporate the Kano model have a business-centric approach. For this research, the term *customer* is replaced by *user*, *student*, or *participant.*

- B (basic requirement)—This is also called a *dissatisfier* or *must-be* factor, as these could be viewed as essential (basic) requirements.
- O (one-dimensional requirement)—The degree of satisfaction has a positive linear correlation with the degree of implementation.
- I (indifferent requirement)—As the name suggests, users do not care about these requirements.
- Q (questionable requirement)—If users judge a feature highly satisfactory for both functional and dysfunctional questions or highly dissatisfactory for both functional and dysfunctional questions, the result is contradictory.
- A (attractive requirement)—A good implementation and performance will greatly increase the customer satisfaction.
- R (reverse requirement)—This factor can be viewed as the inverse of the one-dimensional requirement.

To illustrate the meaning of the Kano model categories covered in the case of an e-learning platform or online course, the following examples are explained for each requirement.

A working website, consistent URLs, and everything viewable on PC and mobile can be considered basic requirements (B). They do not increase the satisfaction if properly included, but would reduce the satisfaction if missing or implemented poorly.

An attractive feature (A) could be a live note-taking function or engaging implementation of gamification (earning points or badges after completing tasks, leaderboards, quests, etc.). Often over time, attractive features become one-dimensional requirements (O) or basic requirements (B).

One-dimensional requirements (O) increase the user's satisfaction proportional to the degree of implementation. Factors that result in a good User Experience (UX) increase the satisfaction proportionally to the degree of implementation. The opposite is also the case as poor implementation of factors that decrease UX also decreases the user's satisfaction.

The Kano model includes states that cannot be captured with other methods in this field. Reverse requirements (R) have the opposite effect of one-dimensional requirements (O) or in some cases of attractive requirements (A), i.e., they decrease (proportionally) the satisfaction when implemented. An example for web-based applications would be pop-up dialogue boxes. This is one of the reasons for choosing this The Kano over others commonly used ones, such as SERVQUAL [12], E-Learning Satisfaction (ELS) [21], and e-SERVQUAL [13]. Dominici and Palumbo also detail two more reasons for this methodology: the first is crucial to this research and allows ex-ante and ex-post analysis, the second is that this model does not assume a linear relationship between the product or service performance and the user satisfaction [5,7].

Questionable requirements (Q) are inconsistencies in the questionnaire answers. If a user answers both highly satisfied if a feature is included and is not included, the answer is unusable, in other words questionable (Q).

2.2 Perceptual Shifts from Before to After the Online Course

This experiment was set up to perform an ex-ante and ex-post analysis. The former equates to a product design study and the latter to a consumption experience study.

The difference from before to after taking the online course is a shift in perception, captured by the Kano model.

This particular experimental setup gave participants anonymized identifiers to later track individual changes. Each individual requirement (by feature and participant) is recorded and compared. These are described as *micro shift* in this paper.

A shift that tracks changes for each feature as a variable. The values of the feature vector are calculated as satisfaction and dissatisfaction indices and is described in Sect. 3.3. These two indices return one numerical value each, their change being referred to as *macro shift*.

3 Methodology

This section covers the experimental setup, questionnaires design, data gathering, analysis, and discussion.

3.1 Experimental Setup and Data Collection

In this study, a total of 16 Japanese students participated, 5 male, 11 female, and 0 other. The average age was 19.82 ($\sigma = 1.70$), ranging from 18 to 24.

This particular system provided by the department of comprehensive psychology offers students to gain additional knowledge and experience by participating in research experiments. Students can earn extra credit by taking part in selected studies. Students can choose in which study they want to participate. All participants are aware of the content of the experiment, the length of the session, and language requirements, amongst other administrative details.

While the whole experiment was scheduled for 90 min, the actual online course was designed to be finished within 60 min. Introduction with an explanation of the setup and concluding remarks with a description of the research goals and implications were provided in Japanese. The online course and following remaining conduct was given in English.

The online course gave students an introduction to depth of field (DoF) in photography. Technical concepts that influence DoF were introduced and artistic use cases were described with examples. Students were quizzed after each section and took a final exam within the online course.

To maximize accessibility and with future research in mind, the online course was set up with the open source website framework *WordPress* and two paid plugins: *LearnDash* and *memberpress*[1]. WordPress has evolved from one of the most popular blogging website to a full-fledged content management system (CMS). It has been used in researching e-learning and online course administration as well [14,15]. The framework allows to create dynamic websites with numerous users at scale. Services include a version that can be self-hosted, i.e., installed and controlled on one's own server. As described in the beginning of the paper, the implementation and management is costly in terms of work effort and expertise. The ability to control open source software, however, is an important advantage for propagation of research and reproducibility.

The LearnDash plugin allows to create complete online courses with integrated enrollment procedure, lectures, topics, quizzes, assignments, and statistics, all managed on WordPress and fully customizable. Similar advantages apply here, as this gives the system administrator full control over the implementation of the online course in functionality and look.

The memberpress plugin allows detailed and customizable management of memberships. This plugin has since been removed from future experimental

[1] The website framework and plugins can be found at the following URLs: https://wordpress.org, https://www.learndash.com, and https://memberpress.com.

setups due to functionality overlap after LearnDash included similar attributes and the former plugin showed unreliable behavior.

Other freely available plugins were used to protect content and create preset user accounts in bulk.

3.2 Questionnaire Design for Kano Model

Questionnaires for the Kano model were given before and after the online course was taken. Each participant received a unique identifier to track the survey results individually while at the same time anonymize the user. Bulk accounts were created in advance and students received login information at the beginning of the experiment.

Twelve factors were collected from 16 participants, each ex-ante and ex-post, resulting in a total of 384 individual Kano requirements formed from the functional and dysfunctional questions. Only one was classified as questionable (Q). This is an indication that the questionnaire is reliably phrased [1].

The following sections detail the data gathered. In addition to the functional and dysfunctional questions necessary for the Kano model evaluation, meta data about students and free form questions were asked as well. The open-ended questions gathered information about students' attitude towards the completion of the online course and if they felt features missing from the questionnaire.

Questions were selected based on previous research, course specific insight, and in accordance with both authors of this paper [2,7,17]. The questions were formatted for readability and professionally proofread.

Ex-Ante Questionnaire (Before). The following information about the participants were collected before the online course: gender (female, male, other) • age • university grade (year) • English language proficiency (none, basic, moderate, advanced, or native level) • knowledge about photography (none basic, moderate, advanced, or professional) • online course features with functional and dysfunctional questions (for Kano analysis) • open ended comment what students wish for in online courses

According to self-assessed English proficiency, 1 student indicated *none*[2], 12 indicated *basic*, and 4 indicated *moderate*.

Ex-Post Questionnaire (After). The following information were collected after completing the online course: overall experience of the course (very bad, bad, neutral, good, very good) • open ended comment what students liked and dislikes about the online course • online course features with functional and dysfunctional questions (for Kano analysis) • open ended comment what students thought was missing in the online course

The last question was asked again to see if participants realized missing features while and after taking the online course.

[2] From several questions written by this student in the open-ended question section, the authors could infer an English comprehension of *basic* rather than *none*.

Features for Kano Model Evaluation. Functional and dysfunctional questions were asked in both the ex-ante and ex-post questionnaire (as described in previous Sect. 3.2 and Sect. 3.2). The following list of 12 features were considered for online courses: User-friendly platform • Certificate of completion • Download of course material • Profile and account page • Quizzes and exercises • Interactive quizzes and exercises • Comment function • Personal tutor • User manual for the platform • Videos • Photos/Graphics • Text

A list with their satisfaction and dissatisfaction indices are shown in Table 2 (Sect. 3.3).

3.3 Results and Discussion

To evaluate the macro shift, a satisfaction index (CS) and dissatisfaction index (CD) was calculated (Eq. (1) and Eq. (2)). Each feature has a pair of CS and CD values.

$$\mathrm{CS} = \frac{\mathtt{A} + \mathtt{O}}{\mathtt{B} + \mathtt{O} + \mathtt{A} + \mathtt{I}} \tag{1}$$

$$\mathrm{CD} = \frac{\mathtt{B} + \mathtt{O}}{\mathtt{B} + \mathtt{O} + \mathtt{A} + \mathtt{I}} \cdot -1 \tag{2}$$

The two equations indicate a ratio to understand a cost-benefit tradeoff of gaining satisfaction and preventing dissatisfaction [4]. Values for CS range from 0 to 1 and higher numbers correlating with higher satisfaction, whereas CD

Table 1. Numerical overview of satisfaction indices (CS_a and CS_p for ex-ante and ex-post respectively), dissatisfaction indices (CD_a and CD_p, ex-ante and ex-post respectively), and standard deviation (σ). Values are rounded to two decimals.

Features	CS_a	CS_p	CD_a	CD_p
User-friendly platform	0.44	0.44	−0.31	−0.31
Certificate of completion	0.19	0.13	0.0	−0.13
Download of course material	0.36	0.25	−0.29	−0.19
Own profile and account page	0.19	0.14	0.0	−0.07
Quizzes and exercises	0.14	0.29	−0.14	−0.14
Interactive quizzes and exercises	0.13	0.2	−0.06	−0.07
Comment function	0.13	0.06	−0.06	0.0
Personal tutor	0.36	0.25	−0.07	0.0
User manual for the platform	0.13	0.44	−0.31	−0.31
Videos	0.44	0.44	0.0	−0.06
Photos/Graphics	0.44	0.56	−0.13	−0.13
Text	0.19	0.4	−0.25	−0.27
σ	0.13	0.16	0.12	0.11

ranges from −1 to 0 and lower numbers correlating with higher dissatisfaction. An overview is given in Table 1.

The correlation ρ of the Customer Satisfaction Index and Customer Dissatisfaction Index before to after changed, becoming less orthogonal and more inversely correlated $corr(\text{CS}_a, \text{CD}_a) = -0.122$ and $corr(\text{CS}_p, \text{CD}_p) = -0.560$.

As the increase in satisfaction is connected with an avoidance of dissatisfaction, a stronger negative correlation points to a clearer response of the user and in turn a system that is easier to optimize.

After calculating the satisfaction and dissatisfaction indices (CS and CD) for all features, their shift from ex-ante to ex-post (ΔCS and ΔCD) is visualized in Fig. 3. Their differences can be inspected in numerically in Table 2.

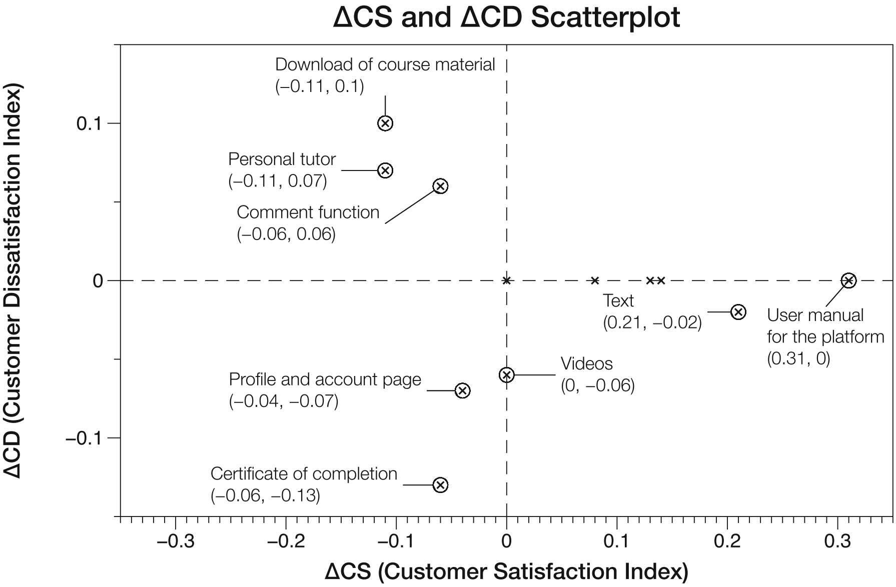


Fig. 3. Ex-ante and ex-post differences of satisfaction and dissatisfaction indices.

Features are labeled for readability, feature on the x-axis from left to right (starting at (0, 0)): *User-friendly platform*, *Interactive quizzes and exercises*, *Photos/Graphics*, and *Quizzes and exercises.*

Macro Shift. With the standard deviations σ of Table 2 relevant changes—mostly for the satisfaction index—can be observed.

Quadrant 1 is empty, indicating for this experiment, there were no features where the CS and CD shift both increased. There were no features that increased in satisfaction and decreased in dissatisfaction at the same time. This can be tied back to the stronger negative correlation from equation $corr(\text{CS}_a, \text{CD}_a)$ to $corr(\text{CS}_p, \text{CD}_p)$, where differences in features fan out from Quadrants 2 to 4.

Quadrant 2 contains three features: *Download of course material*, *Personal tutor*, and *Comment function*. This area indicates a decrease both in satisfaction and dissatisfaction. The users perspective was more positive in the beginning and dropped slightly afterwards. The dissatisfaction ratio also decreased. This quadrant can be interpreted as users' higher expectations being lowered while taking the course. For a concrete example of a personal tutor, the need might be judged more important before starting an online course. After successfully taking the course, a need in actuality seldomly arises. Comment functions are initially perceived more important, possibly as they are ubiquitous in online services.

Quadrant 3 contains two features: *Certificate of completion*, and *Profile and account page*. Their perception was more positive before taking the online course and have potential negative impact when implemented, namely a higher dissatisfaction. Especially the certificate of completion fell into this category. As not all participants could gain a score high enough to gain this certificate, a negative bias could be the reason for this. In future studies, this vector will be compared with a database log of participants who received the certificate.

Quadrant 4 contains only one feature: *Text*. While it indicates an increase in satisfaction, there is potentially increased dissatisfaction as well. For *Text*, however, the shift in ΔCD is close to 0, mainly pointing to an increase in satisfaction. A strong reliance on text for online courses could negatively impact the experience.

Five features show unique locations on the x-axis, having no change in the dissatisfaction index. These features are: *User-friendly platform*, *Interactive quizzes and exercises*, *Photos/Graphics*, *Quizzes and exercises* and *User manual for the platform*. Especially the feature *User-friendly platform* shows no change but has a high CS and low CD, which needs a closer look using the micro shift discussed in paragraphs below. The need for a user manual can be seen from the graph. For short-term experiments, this can be viewed as more important and will be compared with long-term experiments in future (ongoing) research. Kind of quizzes had less influence on the satisfaction rate. The question about interactive quizzes will be dropped for future research. There was also no big difference in implementation of photos or graphics. Both had a strong satisfaction index, which increased after taking the course, with no change for the dissatisfaction index.

One feature shows a unique location on the y-axis, with no change in the satisfaction index: *Videos*. Videos are usually deemed attractive features, which is reflected by the high CS_a and CS_p. A slight dip in dissatisfaction from before to after the course might be a degree of sufficiency of other multimedia.

Micro Shift. To evaluate the micro shift, each of the Kano model cells (16 participants with 12 features each) were contrasted ex-ante and ex-post. Results are discussed for features, where the micro shift indicated additional insight.

The feature *Download of course material* was the only one who showed a micro change in two places from reverse to indifferent `R` → `I`. At the same

Table 2. Numerical overview of the changes of satisfaction, dissatisfaction indices (ΔCS and ΔCD) from before to after the online course, and standard deviation (σ). Values are rounded to two decimals.

Features	ΔCS	ΔCD
User-friendly platform	0.00	0.00
Certificate of completion	−0.06	−0.13
Download of course material	−0.11	0.10
Own profile and account page	−0.04	−0.07
Quizzes and exercises	0.14	0.00
Interactive quizzes and exercises	0.08	−0.00
Comment function	−0.06	0.06
Personal tutor	−0.11	0.07
User manual for the platform	0.31	0.00
Videos	0.00	−0.06
Photos/Graphics	0.13	0.00
Text	0.21	−0.02
σ	0.13	0.06

time, two values each were changed from attractive and basic both to indifferent (`A` → `I` and `B` → `I`). This explains the shift of ΔCS and ΔCD into quadrant 2.

For the inclusion of *Personal tutor*, three users shifted from attractive requirement to indifferent requirement (`A` → `I`), the highest number in a feature vector. Two users, however, shifted to the opposite (`I` → `A`). Although the macro shift is similar to the feature *Download of course material*, the implications are different. Three users expected to need a tutor did not after taking the online course, where two rated the feature as attractive. This could be due to how tutors are anticipated, but not missed afterwards.

Videos had an interesting shift as it had the highest number of attractive requirement, but also the highest number of change from attractive to indifferent (three, the same as *Personal tutor*). This can be interpreted as high expectation with a contrast to the consumption experience. There could be a bias in this short online course, that relied on photographs and graphics. Further research is necessary to see any long-term effects on this feature.

The feature *Text* gained the most attractive requirement values, three `I` → `A` and one `O` → `A`. One user changed from attractive to one-dimensional (`A` → `O`). The short online course could present a bias here as well, as formatting, formulation, and other factors of text material could produce different outcomes in courses spanning weeks or months.

The availability of *User manual for the platform* gained as well, which corresponds with open-ended survey answers that students wished for a Japanese translation or more explanation on how to use the online course system. Although two values were changed from basic to one-dimensional (B → O), two shifts in I → A and one in O → A.

Comments of students helped to interpret the results and shifts. A common positive theme was regarding the interest in the online course content and the visual appeal. The ability to revisit the content and re-take quizzes was another positive note. Negative comments were related to English comprehension and lack of familiarity with the online course system.

3.4 Lessons Learned

From the results analyzed and visualized in Fig. 3, RQ1 can be confirmed, as the Kano model can help identify a shift in expectations versus consumption experience. With the help of plotting the results or analyzing pair-wise ΔCS and ΔCD, the satisfaction and dissatisfaction indices could be classified, affirming RQ2.

RQ3 asked the additional insight of requirements with low to no ΔCS or ΔCD. The previous section Micro Shift lists such examples and reinforces the value of using both macro and micro shifts for this research.

The short-term online course used in this study revealed that being unfamiliar with the online system creates some unease in students. Regardless of the setup being in accordance with best practices known from online course design, users need a period of adaptation to get used to new platforms. A short introduction video rather than a brief slideshow might increase students' readiness. Such a "how-to" page including a short video should be accessible to students throughout the course.

After students completed the online course, their perception towards material that can be downloaded from such a system changed to overall indifferent (noticeable in the micro shift, which the macro shift did not indicate). This requirement ranged from reverse to attractive before taking the course, showing that students have various expectations towards this requirement, which turned out mute after completion.

3.5 Limitations

This research was conducted with 16 participants with a majority of female students. All participants were Japanese and about 20 years old. The homogeneity of Japanese society could allow small sample sizes, but generally this is a limitation of this study.

The online course could be taken in one session (within 60 min). Longer use of an online platform could lead to different outcome. Long-term and short-term prioritization of features are under investigation by the authors in ongoing research.

Students who took the course had moderate English comprehension, which could influence the questionnaire outcome. A followup experiment is planned with the same system setup translated to Japanese.

RQ3 needed the support of free-form questionnaires to identify causality of student's micro shifts. Thinking of at-scale use of macro and micro shift, natural language processing (NLP) tools such as large language models could be used to automate additional insights.

4 Conclusions

This research investigated the macro and micro shifts in the perception of online course features using the Kano model. Contrasting the expectations of users before taking a course with the consumption experience after completion provides useful insight to online course features in how their implementation would influence student satisfaction. While an ex-ante and ex-post comparison provides a powerful method for user satisfaction research, this paper shows that additional insight can be gained by looking closer at requirements that do not show a significant change from before to after.

For online course design, the perception and change thereof should be interpreted and taken into consideration for learning and teaching objectives. Where the *macro shift* can show general tendencies for student satisfaction, the *micro shift* can reveal teaching necessities being at odds with a learner's inclination. The ex-ante and ex-post analysis showed significant shifts in how users perceive online course features and is expected to lead to further insights in comparison with long-term experiments.

For future research, larger scale and long-term comparisons are aimed to allow comparisons across different teaching subjects. Different course topics are planned to be investigated with students from different study fields.

Acknowledgements. This research was supported by JSPS KAKENHI Grant Number 22K02874.

References

1. Basfirinci, C., Mitra, A.: A cross cultural investigation of airlines service quality through integration of servqual and the kano model. J. Air Transp. Manag. **42**, 239–248 (2015). https://doi.org/10.1016/j.jairtraman.2014.11.005
2. Bearden, W.O., Netemeyer, R.G., Haws, K.L. (eds.): Handbook of Marketing Scales - Multi-Item Measures for Marketing and Consumer Behavior Research. SAGE Publishing, New York (2010)
3. Becher, T.: The significance of disciplinary differences. Stud. High. Educ. **19**(2), 151–161 (1994). https://doi.org/10.1080/03075079412331382007
4. Berger, C., Blauth, R.E., Boger, D.: Kano's methods for understanding customer-defined quality. Cent. Qual. Manag. J. **2**, 3–36 (1993)
5. Chaudha, A., Jain, R., Singh, A.R., Mishra, P.K.: Integration of kano's model into quality function deployment (QFD). Int. J. Adv. Manuf. Technol. **53**(5), 689–698 (2011). https://doi.org/10.1007/s00170-010-2867-0
6. Chen, L.H., Kuo, Y.F.: Understanding e-learning service quality of a commercial bank by using kano's model. Total Qual. Manag. Bus. Excellence **22**(1), 99–116 (2011). https://doi.org/10.1080/14783363.2010.532345

7. Dominici, G., Palumbo, F.: How to build an e-learning product: factors for student/customer satisfaction. Bus. Horiz. **56**(1), 87–96 (2013). https://doi.org/10.1016/j.bushor.2012.09.011
8. Kano, N., Seraku, N., Takahashi, F., ichi Tsuji, S.: Attractive quality and must-be quality. J. Jpn. Soc. Qual. Control **14**(2), 147–156 (1984). https://doi.org/10.20684/quality.14.2_147
9. Katai, Z.: Promoting computational thinking of both sciences- and humanities-oriented students: an instructional and motivational design perspective. Educ. Tech. Res. Dev. **68**(5), 2239–2261 (2020). https://doi.org/10.1007/s11423-020-09766-5
10. Chen, L.H., Lin, H. C.: Integrating kano's model into e-learning satisfaction. In: 2007 IEEE International Conference on Industrial Engineering and Engineering Management, pp. 297–301 (2007). https://doi.org/10.1109/IEEM.2007.4419199
11. Mikulić, J., Prebežac, D.: A critical review of techniques for classifying quality attributes in the kano model. Manag. Serv. Qual.: Int. J. **21**(1), 46–66 (2011). https://doi.org/10.1108/09604521111100243
12. Parasuraman, A., Zeithaml, V.A., Berry, L.: Servqual: a multiple-item scale for measuring consumer perceptions of service quality. J. Retail. **64**, 12–40 (1988)
13. Parasuraman, A., Zeithaml, V.A., Malhotra, A.: ES-QUAL: a multiple-item scale for assessing electronic service quality. J. Serv. Res. **7**(3), 213–233 (2005). https://doi.org/10.1177/1094670504271156
14. Quesenberry, K.A., Saewitz, D., Kantrowitz, S.: Blogging in the classroom: using wordpress blogs with buddypress plugin as a learning tool. J. Advertising Educ. **18**(2), 5–17 (2014). https://doi.org/10.1177/109804821401800203
15. Rodgers, A.R., Puterbaugh, M.: Digital badges and library instructional programs: academic library case study. J. Electron. Resour. Librariansh. **29**(4), 236–244 (2017). https://doi.org/10.1080/1941126X.2017.1378542
16. Selim, H.M.: Critical success factors for e-learning acceptance: confirmatory factor models. Comput. Educ. **49**(2), 396–413 (2007). https://doi.org/10.1016/j.compedu.2005.09.004
17. Selvi, K.: Motivating factors in online courses. Proc.-Soc. Behav. Sci. **2**(2), 819–824 (2010). https://doi.org/10.1016/j.sbspro.2010.03.110
18. Sun, P.C., Tsai, R.J., Finger, G., Chen, Y.Y., Yeh, D.: What drives a successful e-learning? an empirical investigation of the critical factors influencing learner satisfaction. Comput. Educ. **50**(4), 1183–1202 (2008). https://doi.org/10.1016/j.compedu.2006.11.007
19. Violante, M.G., Vezzetti, E.: Virtual interactive e-learning application: an evaluation of the student satisfaction. Comput. Appl. Eng. Educ. **23**(1), 72–91 (2015). https://doi.org/10.1002/cae.21580
20. Wang, Y.S., Bauk, S., Šćepanović, S., Kopp, M.: Estimating students' satisfaction with web based learning system in blended learning environment. Education Research International 2014 (2014). https://doi.org/10.1155/2014/731720
21. Wang, Y.S., Wang, H.Y., Shee, D.Y.: Measuring e-learning systems success in an organizational context: scale development and validation. Comput. Hum. Behav. **23**(4), 1792–1808 (2007). https://doi.org/10.1016/j.chb.2005.10.006

Glow-Mind: An Input/Output Web System for Sharing Feelings by Pressing a Button

Kanan Abe(✉), Taai Tsukidate, Yo Kuwamiya, Hiroki Echigo, and Minoru Kobayashi

Meiji University, 4-21-1 Nakano, Nakano-Ku, Tokyo, Japan
{kanan.abe,taai.tsukidate,yo.kuwamiya,hiroki.echigo}@koblab.org, minoru@acm.org

Abstract. Uncomfortable silences can disrupt meetings and conferences. One of the reasons for these silences may be that participants are unable to share their intentions at the meeting, which makes it difficult to decide whether discussions should move on or go into greater depth. This study proposes a method for visualizing feelings and conveying them to participants in a meeting. Our design *Glow-mind* involves button systems in web browsers which can share feelings anonymously. This paper introduces the system requirements, tests the system, and sets out the results. To facilitate real-life meetings, we also consider a new user interface in which the size of each button is determined by the number of times it is pressed.

Keywords: Meeting support · Face-to-face communication · Teleconference · Button

1 Introduction

Discussions are essential in organizations and groups to further mutual understanding. There are many opportunities for discussion in companies and schools, for example, and discussions need to be active so that meetings within a limited timeframe can be meaningful.

However, uncomfortable silences may mean that discussions do not proceed smoothly. Participants in the meeting might infer something from the silence and take action to resolve the situation if they consider the silence inappropriate. Silences should therefore be avoided as they can be troublesome for the participants. We hypothesize that one reason for these silences could be that the feelings of the participants are not made clear, so that it is difficult to determine whether discussions should move on or go into greater depth.

To facilitate meetings, this study proposes a button system called *Glow-mind,* which works in web browsers. *Glow-mind* visualizes participants' feelings anonymously when they press buttons during discussions. The study involves feelings such as "Agree", "Disagree", "Have an opinion", "I see", "Good" and "Don't know".

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 126–140, 2022.
https://doi.org/10.1007/978-3-031-20218-6_9

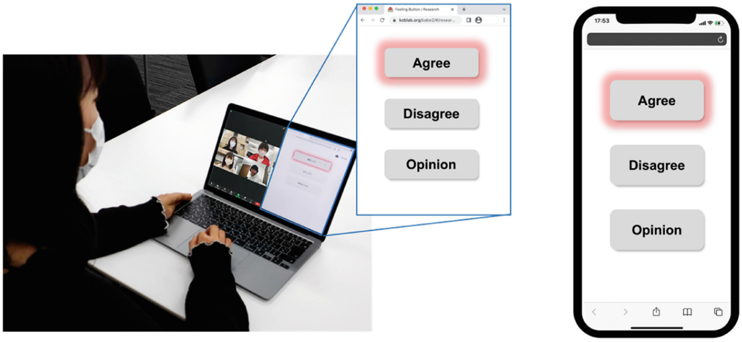


Fig. 1. Using the system.

The study covers not only face-to-face meetings but also teleconferences. The system limitations of teleconferences make it difficult for multiple people to speak at once, and to make eye contact or blink. They therefore receive less information than in face-to-face meetings, which makes active discussion more difficult. We therefore propose a system for teleconferences. Figure 1 shows an image of the system in use.

The contributions of this paper are as follows:

- We introduce our previous studies on sharing participants' feelings to facilitate meetings.
- We present the design features of *Glow-mind*, which supports participants in sharing their feelings during meetings.

2 Related Work

2.1 The Meanings of Silence

Theorists argue that silence has a variety of meanings. Chowdhury et al. [2] investigated the function of silence in dialogs, clustering the results. They found that silence had certain meanings such as preparing to respond, hesitating or considering asking a question. Equally, in an analysis of decision-making meetings, Kurosu et al. [6] found that some participants said very little.

We believe that the polysemy of silence obscures its meaning, and hinders the smooth progress of discussions. The purpose of our study is to visualize the function of silences in terms of making meetings meaningful.

2.2 Support for Sharing Feelings with Biosensor Data

Snyder et al. [11] developed MoodLight, an interactive ambient lighting system. It responds in real time to skin potentials, which are biosensor data related to an individual's level of arousal.

Howell et al. [5] developed Hint as a communication trigger, a system that uses changing clothing patterns in response to emotions inferred from biometric signals.

However, visualizing feelings with biosensor data potentially involves unintended participant output and privacy issues. Therefore, we propose a system in which participants themselves input the feelings they wish to be visualized.

2.3 Support for Sharing Feelings with Participants' Input

Apart from our own approach, other methods use participants' input. gIBIS, developed by Conklin et al. [3], supports discussion by visualizing the position and meaning of opinions under discussion (agree, disagree, support, questions, etc.). On the other hand, direct expression of feelings is not favored in Japanese culture. Therefore, *Glow-mind* provides a solution for the Japanese context by supporting the visualization of feelings that are difficult to express directly. This aims to facilitate the progress of meetings and make them meaningful.

PICALA, a system developed by Yumura et al. [14], visualizes the feelings of the audience at conference presentations using the color of the lighting. The input is a PC or smartphone used by the audience, and the output is the lighting at the venue. However, this method involves setup costs and associated costs for lighting, whereas our system requires only common devices such as PCs and smartphones. In addition, PICALA uses a separate device for input and output. We believe that this method can cause strain and a loss of focus, as participants need to look at both the input device and the output. Hence, we propose a system where input and output involve a single device.

Naruhodo Button, a physical button system developed by Yoshida et al. [13], enables participants to give positive feedback by using sounds expressly or casually in face-to-face brainstorming. A common aspect between this study and our own is that both share feelings by pressing buttons, but they are differentiated by the method of visualizing the feelings. We consider that the use of sounds is likely to disrupt meetings. We therefore propose using glowing buttons rather than sound. Another difference between this study and our own involves anonymity, and whether or not the system is physical.

In a similar way to our study, Suzuki et al. [12] developed FeelLight for research using a single input/output device. FeelLight uses an optical button device to communicate "some sort of intention" to a remote person. Our own study is more specific in its communicative intent.

Reaction features like emojis and polling in video-conferencing systems such as Zoom [17] and Microsoft Teams [7] are examples of ways in which participants' input is used. Our study differs from the reaction feature in that it involves different media, and the user is not identified. In other words, *Glow-mind* visualizes feelings anonymously. Using text rather than emojis is unique in terms of visualizing participants' feelings. The difference between the polling feature and our system is the timing of feedback. While the polling feature requires the meeting to pause in order to share the results, our system allows people to share their feelings in real time, as the meeting progresses.

3 System Design

3.1 Requirements

We propose a button web system called *Glow-mind* to help users visualize feelings and to facilitate meetings. The requirements for our system are as follows.

R1 Input/output is available in a single system.
R2 Real-time communication is possible.
R3 The output results are unique.
R4 Anonymity is preserved.
R5 Input methods are unified.

The aim of R1 is to reduce the number of locations to which participants need to pay attention. The system allows input by pressing, and output follows when the device lights up.

The intention of R2 is to synchronize the system with the progress of the meeting. This is because the aim is to visualize the feelings that influence the progress of meetings. To meet the objectives of R2, our system operates online.

R3 hopes to prevent the possibility of participants reading the output result in different ways. Having to think about the meaning of the result means they could lose their concentration in a discussion. Therefore text, rather than emojis or emoticons, meets the objectives of R3.

We established R4 with the option of visualizing negative or critical feelings. These can be challenging to express in words, so R4 is satisfied by allowing the participant to remain anonymous. Participants cannot attribute the feelings visualized by the system to a particular person. The system is nevertheless capable of detecting participants without using their real names.

We designed R5 to avoid differences in the operating method of each button. The buttons are therefore parallel and are unified into a single click to avoid confusion.

3.2 Our Previous Studies on the Research Question

Our study addresses ways of overcoming silence, which can otherwise prevent meetings from being meaningful. We propose *Glow-mind*, a button system. It allows participants to visualize their feelings, improving their satisfaction with a meeting.

Up to now, we have developed four versions of the system. These systems are named Ver. 1 through Ver. 4. Ver. 1 is a concept-verification system for visualizing feelings. Ver. 2 is based on the results of a survey about feelings a person might wish to show or detect. Ver. 3 can set out the visualized feelings in text form. Ver. 4 allows participants to choose from 12 pre-installed buttons based on how useful they think they are to express their feelings. This version is designed for an experimental purpose to investigate the participants' preferred settings for expressing their feelings by letting them freely choose the buttons from a number of options. This feature would lead to participants experiencing a high cognitive load when selecting the buttons from 12 choices while following the meeting in progress. Thus, in the future development of a practical system, we are considering providing pre-defined sets of button groups that meet typical meeting purposes. Figure 2 shows the screen images, and Table 1 shows the features of each version.

The following sections present our previous studies on four research questions:

- System configuration
- Types of feeling to be visualized

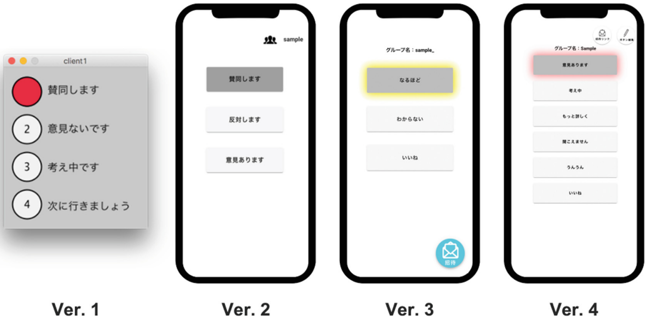


Fig. 2. *Glow-mind* screens from our previous studies, where the top button in Japanese is pressed. Ver. 1 has four predetermined buttons: "Agree", "No opinion", "Thinking" and "Next topic". The buttons in Ver. 2-4 are set freely by participants.

Table 1. The features of the application in our previous studies.

	Devices	Setup method and number of buttons	Color when you press		Color when others press	
			Button	Surround	Button	Surround
Ver. 1	PC	Predetermined by experimenter Set 4 before use	Red	–	Blue	–
Ver. 2	PC, smartphone, tablet	Predetermined by experimenter Set 3 before use	Gray	–	–	Red
Ver. 3	PC, smartphone, tablet	Predetermined by a participant Set 3 before use with text	Gray	Any color	–	Any color
Ver. 4	PC, smartphone, tablet	Set by participants at any time Set 0 to 12 from choices	Gray	9 colors	–	9 colors

- Feedback method
- Showing the log of buttons pressed

System Configuration

We took different devices into consideration for the system. Ver. 1, implemented in Processing [9], runs only on PCs. This version was not designed for use in teleconferences, as it worked only in the same network environment. Additionally, it required participants to download the program. This meant that the system configuration was not adapted for many people to use. Based on these issues, we set the following two requirements in addition to the five previously listed.

R6 Available between remote locations
R7 Ease of installation

To meet these requirements in Ver. 2 and in later versions, we implemented the system in Vue.js [16] and Google Firebase [4]. When a participant presses a button, data on who pressed which button and when are routed to the database via the server (Fig. 3). This makes the system accessible in the browser, eliminating the need for program distribution. Moreover, the system is available not only on PCs but also on smartphones and tablets. To enhance this further R7, we designed the system to allow participants to start without signing in, simply by accessing an URL or reading a QR code [10]. These specifications have made it possible for participants to use the system easily, even when they are located remotely from one another and are using different devices.

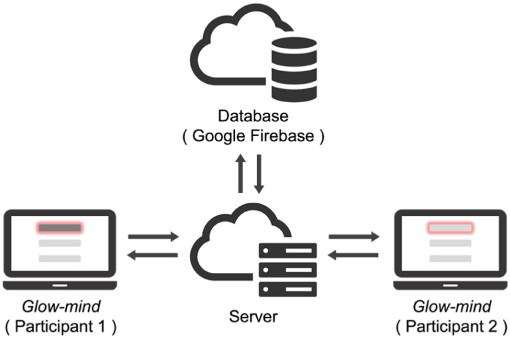


Fig. 3. System overview

Types of Feeling to Be Visualized

Related studies suggest that participants' silence can represent a range of different emotions. We therefore explored the types of feeling that would need to be visualized on the button system.

We provided Ver. 1 with four feelings which could affect the progress of a meeting, based on a study that analyzed the reasons for silence [6]: 1) Agree, 2) Disagree, 3) Thinking and 4) Next topic. We considered four reasons why participants might not express their feelings: 1) they are not opposed to what is being said; 2) they are reluctant participants; 3) they would not dare to say "I am thinking"; 4) they are giving the impression that they are bored, and are wanting the meeting to end. Our preliminary face-to-face experiments using Ver. 1 showed that some buttons were rarely used, or that the timing involved in pressing them was ambiguous. We found it necessary to explore further which feelings to visualize.

We therefore surveyed the feelings that conference participants wished to express or perceive. As a result of the survey, we included three buttons in Ver. 2 of our system and evaluated them: “Agree”, “Disagree” and “Have an opinion”. The feelings ranked in the top three of both the “wish to express” and “wish to perceive” question items. We conducted an experiment in which participants used Ver. 2 during online meetings. We counted the number of buttons pressed and asked for feedback from participants. These data showed that the actual needs captured in the experiment were different from those in the presurvey.

Following the feedback we developed Ver. 3, in which the participants themselves set how their three feelings would be visualized via *Glow-mind* in any text before the start of the meeting. In user tests with Ver. 3, some participants commented, “It was hard to decide on the feelings before meetings because we did not know which feelings we would want to share. In addition, considerable freedom of choice made it difficult to set the feelings.” We also received requests to try different feelings to visualize other feelings than the three which they set. The feedback indicated that the number of buttons in the system was an important issue to consider.

A field experiment using Ver. 4 is currently underway to examine suitable types of feeling and how many would be appropriate. The experiment involves participants using *Glow-mind* in real meetings both face-to-face and online. Ver. 4 allows users to set the type and number of feelings freely, at any time and as many times as they wish. We set 12 options for feelings based on related works and comments received in experiments. Table 2 shows the 12 options and their classification.

Table 2. Feelings and classifications in system options.

Classification	Example
Opinion	Agree
	Disagree
	Have an opinion
	Have no opinion
Confirmation of progress of discussion	On to the next topic
	Thinking
Accelerating discussion	Uh-huh
	I see
	Good
Exploration of discussion	More details
	Don’t know
Environmental improvement	Can’t hear you

Feedback Method

We also investigated the feedback method. In Ver. 2, a button turns gray when it is pressed. When someone else presses a button, the color around the button turns red. In an experiment using Ver. 2, one participant said, “My emotions were not linked to the color of the button turning gray". In response to this comment, we developed Ver. 3 so that when the button is pressed, it glows just as the other participants see it.

Additionally, a participant using Ver. 2 suggested, "You should make the color of the glow more colorful, because sometimes I don't notice that the button is glowing if I am involved in conversation". Hence, Ver. 3 allows users to set the color around the buttons, with a choice of nine: red, yellow, green, blue, sky blue, pink, orange, purple and brown. These are based on the Model Color Palette for the Color Universal Design GUIDE BOOK [8]. The specification means that participants can identify the button which is glowing by color, even when the buttons are hard to see.

Showing a Log of Buttons Pressed

The buttons in Versions 1 to 3 only glowed for a short time, for 1s in Ver. 1, and 5s. in Ver. 2-4. Furthermore, *Glow-mind* does not show a log of buttons pressed. Users had to look at the system within 5 s of a button being pressed to notice that someone had visualized their feelings. Some users commented that they felt pressurized to make sure they did not miss someone pressing a button. To resolve the issue, this paper proposes a method where the size of each button represents the number of buttons pressed. Section 4 will explain the details.

3.3 System Operation

Our system involves buttons with bidirectional input/output, and operates in a web browser. Figure 4 shows the screens when the system boots up, when a participant presses a button, and when other participants press a button.

In our previous study [1], we conducted four controlled experiments in which groups of four participants had online discussions for 30 min. Table 3 shows the results of evaluating the system. All 16 participants indicated that the design of the buttons in the system made them easy or very easy to push. We also found significant differences between the cases which used our system and those which did not. These involved two aspects: "active discussion" and "communication between participants".

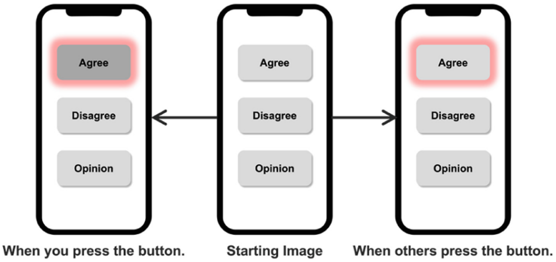


Fig. 4. How the system operates

Table 3. Results of the online experiment using Ver. 2 of *Glow-mind*.

	Question items	Avg
Design	Were the buttons designed so that they were easy to push? (−2: Very hard to push, to +2: Very easy to push)	1.625
Feedback method	Were you hesitant about pushing the buttons? (−2: Not hesitant at all, to +2: Very hesitant)	−0.75
	Did being anonymous affect the likelihood of your pressing the button? (−2: No impact, to +2: Very strong impact)	0.5
	Did the gray color of the buttons enable you to indicate the feelings you wished to express? (−2: I couldn't, to +2: I could)	0.56
	Did the red color around the buttons enable you to read other participants' feelings? (−2: I couldn't, to +2: I could)	1.31
Reactions	Did you notice who pressed the button? (−2: Not at all, to +2: Very well)	−1.125
	Did you feel that other participants knew you had pressed the button? (−2: Not at all, to +2: Very well)	−0.625
Impression	How did you feel about the buttons? (−2: Uncomfortable, to +2: Comfortable)	0.69

4 User Interface Which Displays Logs

Previous studies indicated that our system needed a log of buttons pressed. This would allow participants to see which buttons had been pressed while they were not looking.

One way of displaying logs involves a live stream on YouTube [15] with text. However, contrary to our requirements R1, this method has separate chat input and output locations. Thus, this paper proposes to represent the log using the size of the buttons. It allows participants to recognize the number of buttons pressed intuitively.

4.1 Ideas for Arranging and Adjusting the Buttons

This section will explain ideas for arranging and adjusting the buttons. The button size is calculated from 100 pieces of data from buttons pressed. The following presents some ideas for arranging the buttons and adjusting them to different sizes. Table 4 shows an example of data from buttons pressed. Figure 5 represents two ideas for the user interface using the data.

Figure 5-a arranges the buttons vertically, as in Versions 1 to 4. The height of each button is calculated from the number of buttons pressed. However, it falls short of indicating the logs if all 12 buttons are active, because 12 buttons do not fit on a smartphone screen (Fig. 6). It is very important that they fit on a screen without the user needing to scroll up and down, since the buttons also serve as output. Therefore, the vertical layout is unsuitable for this method.

Figure 5-b arranges the buttons in two rows, and is known as a masonry layout. The ratio of the number of buttons pressed determines the height of each button. The next section indicates the procedure for calculating the size of the buttons.

Table 4. An example of button data.

Name of button	Press count in the last 100
I see	50
Agree	25
Disagree	15
Uh-huh	10
Good	0

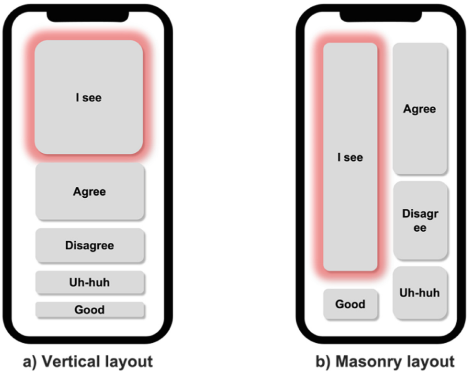


Fig. 5. Two ideas for arranging and adjusting the user interface

4.2 Calculating the Height of Buttons

This section explains how the height of the buttons and the glow around the buttons is calculated according to the method in Fig. 5-b. The size of each button is calculated from the data from the last 100 times the button was pressed, as shown in Table 4. Figure 7 shows an interface design of each button where *n* is the number of buttons set by the user from 1 to 12.

The width of the buttons including the glow area is half the width of the screen, so that the buttons can be arranged in two rows. The formula for the height of the button area and glow area (H_n) and the height of button (B_n) are:

$$H_n = 2h\frac{d_n}{100} \qquad \text{(F1)}$$

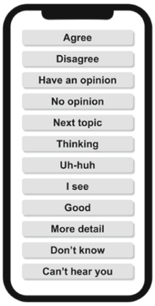


Fig. 6. Screen with all 12 buttons active in a vertical layout.

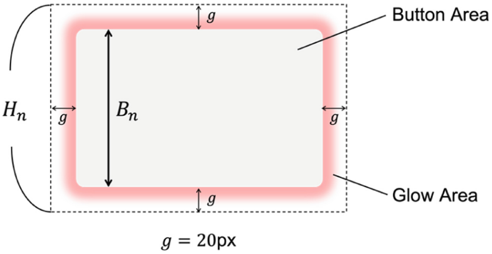


Fig. 7. Design of each button.

$$B_n = 2h\frac{d_n}{100} - 2g \tag{F2}$$

where h is the height of the displayable area, g which is the height and the width of the glow area is set at 20px. d_n is the data from each button pressed. It should be noted the default value of d_n is not zero, but is:

$$d_n = \frac{100}{n} \tag{F3}$$

4.3 Limitation

We set g to 20px to clearly show the button pressed to users. However, we are not sure whether the value is appropriate and will discuss it further.

Our system can change the type of button during use. If that happens, d_n automatically resets to the default (formula (F3)). We consider this processing has the effect of users noticing that the button type has changed.

The buttons may protrude from the screen when using formula (F1) to calculate the height of the button area and glow area (H_n). In that case, the system adjusts the placement and height. This process does not have problems since our system does not show the exact number of times each button is pressed but is used to visualize participants' feelings.

In the method of expressing the log by the size of the buttons as described above, when the size of the buttons changes, the position of the buttons changes, which may lead to the wrong button being pressed. Therefore, the system should be implemented so that the position of the buttons does not change significantly.

5 Discussion

We have proposed a means for meeting participants to visualize their feelings and to facilitate meetings. Meetings take place in a wide variety of forms depending on the number of people, the purpose of the meeting and whether or not it is face-to-face. Our system can be adapted to each type of meeting and the devices used.

The number of participants is a major factor defining the meeting's characteristics. In terms of using our system, previous studies have shown that in trinitarian meetings, participants often notice who has pressed the button. In other words, the requirement R4 (anonymity) mentioned in Sect. 3.1 is hard to establish when using our system in small meetings. We need to perform further experiments to confirm the effect of using our system in meetings with different numbers of participants.

The purpose of the meeting is another major factor influencing meeting characteristics. Each meeting has a general aim, such as decision-making, liaison or brainstorming. Participants are likely to want to visualize their feelings in different ways depending on the purpose and phase of a meeting. For instance, brainstorming involves phases of divergent thinking and convergent thinking. In the divergent thinking phase, participants will want to visualize "I see", "Uh-huh" and "On to the next topic" to encourage ideas. In the convergent thinking phase, on the other hand, they will want to visualize "More details", "I don't know" and "Agree" to bring their ideas into shape. Our system supports various kinds of meeting purposes, but different sets of buttons need to be designed to accommodate each meeting need.

Recently, online meetings have started to be used daily. Whether or not a meeting is face-to-face is another major factor we have to consider. As our system works with a web browser, it can be used on various devices, including smartphones, tablets, and PCs. Laptop PCs are powerful tools for keeping records or viewing documents in meetings, whether they are face-to-face or not. When using laptop PCs for such purposes, using our system on smartphones simultaneously is good; users can avoid switching between the windows of our system and other applications, and can express their feelings by just tapping on the phone. However, when participating in face-to-face meetings without using laptop PCs, people are reluctant to use smartphones because they may appear not to be focused on the meeting. In such cases, using our system on laptop PCs looks better and makes them feel at ease.

Users of *Glow-mind* may sometimes forget to press buttons at meetings. We do not think that is such a big issue, because forgetting to press the button may be thought to indicate that the participant does not need to express their feelings using the system. On the other hand, it is a problem if a participant often misses seeing the others' feelings displayed on our system, because if the display is not seen, our system has no way to affect the users. This often happens when our system's window is hidden by other windows. One way to avoid this happening is to use our system on a separate display or to use a smartphone, but to make the system work effectively a way of attracting the users' attention should also be built into the system.

6 Future Work

6.1 Field Experiment

As noted in Sect. 3.3, in our previous experiments the system contributed to "active discussion" and "communication between participants". Furthermore, all participants said the design of the button system made it easy to press buttons. However, the experiments lasted only one day, yielding short-term results. Moreover, we presupposed that participants were using the system proactively, but whether they were using the system voluntarily was uncertain. To ascertain whether this is the case, the system will be trialed in field experiments involving real-life meetings, both face-to-face and online. These will also take place over an extended period of time. The field experiments will measure to what extent the participants are monitoring the output of the system and changing their behavior in the meeting based on that.

6.2 Types of Feeling to Visualize

As previously stated in Sect. 5, the feelings it is useful to visualize depend on the purpose or phase of a meeting. A field experiment with Ver. 4 will study which feelings to visualize in each meeting environment. Participants will be able to set the types of feeling freely, and the range of feelings available for each meeting environment will be adjusted on the basis of the results. We will develop Ver. 5 in a masonry layout using the results.

6.3 Displaying Logs

As noted in Sect. 4.3, the method of displaying logs by changing the size of buttons may lead to the wrong button being pressed. To avoid this risk, we have a completely different method from the one proposed in Sect. 4. Notably, the more frequently used buttons are hotter, and the less frequently used buttons are cooler, expressed by the size of the Glow Area (Fig. 7), glowing seconds and color intensity of the glow around the buttons. We will carefully consider a suitable log expression method by comparing the methods of expressing the temperature of the buttons (Sect. 6.3) and changing the size of the buttons (Sect. 4).

6.4 Evaluation of the Discussion

In addition to field experiments, we plan to conduct controlled experiments to examine the extent to which the system facilitates meetings. Eight evaluation items are shown below, and these will be assessed through a survey and an analysis of recordings.

a. Short periods of silence.
b. Reaching a conclusion in a short time.
c. Wide perspective: the number of ideas.
d. Participants are satisfied with the process.
e. Participants are satisfied with the conclusion.
f. Good communication between participants.
g. Participants are able to focus on the discussion.
h. Participants achieve the aim of the meeting.

7 Conclusion

This study has aimed to resolve the issue of silence in meetings, which can prevent the meeting proceeding smoothly. We hypothesized that the problem was caused by the fact that participants were unable to share their thoughts, and that solving the problem would lead to more meaningful discussions. We therefore proposed *Glow-mind*, a button system using browsers for decision-making meetings. Participants can use our system to visualize their feelings while remaining anonymous at all times. This paper introduced seven requirements for the system.

Significant differences have already been identified between meetings where our system is used and those where it is not. These involve two aspects: "active discussion" and "communication between participants". A previous experiment also showed that the design of the button system makes it easy to press.

A field experiment is currently underway using the system we developed to examine the types of feeling which can best be visualized. This paper has proposed a method for showing a log of buttons pressed in real-life meetings. The method involves determining the size of each button by using the number of buttons pressed. Further experiments will be conducted using this system to verify how effective it is in facilitating meetings.

Acknowledgments. This work was supported by JSPS KAKENHI Grant Number JP18K11410.

References

1. Abe, K., Tsukidate, T., Kuwamiya, Y., Kobayashi, M.: Designing a state of mind visualization button to facilitate meetings. In: DICOMO 2021, pp. 774–783 (2021). (in Japanese)
2. Chowdhury, S.A., Stepanov, E., Danieli, M., Riccardi, G.: Functions of silences towards information flow in spoken conversation. In: Proceedings of the Workshop on Speech-Centric Natural Language Processing, pp. 1–9 (2017)
3. Conklin, J., Begeman, M.L.: gIBIS: a hypertext tool for exploratory policy discussion. ACM Trans. Inf. Syst. **6**(4), 303–331 (1988)
4. Google Firebase. https://firebase.google.com/. Accessed 01 July 2022
5. Howell, N., et al.: Biosignals as social cues: ambiguity and emotional interpretation in social displays of skin conductance. In: DIS 2016, pp. 865–870 (2016)
6. Kurosu, M., Yamadera, H., Mimura, I., Sumino, S.: The analysis of the actual meeting (1). How the groupware system can support the group process. In: Groupware Research Group, vol. 1995, no. 38, pp. 25–30 (1995). (in Japanese)
7. Microsoft Teams. https://www.microsoft.com/en-us/microsoft-teams/group-chat-software. Accessed 01 July 2022
8. Model Color Palette for the Color Universal Design GUIDE BOOK. https://www3.dic-global.com/dic-graphics/navi/color/pdf/cud_guidebook.pdf. Accessed 01 July 2022. (in Japanese)
9. Processing. https://processing.org/. Accessed 13 June 2022
10. QR Code. https://www.denso-wave.com/en/adcd/fundamental/2dcode/qrc/index.html. Accessed 01 July 2022
11. Snyder, J., et al.: MoodLight: exploring personal and social implications of ambient display of biosensor data. In: CSCW 2015, pp. 143–153 (2015)
12. Suzuki, K., Hashimoto, S.: FeelLight: a communication device for distant nonverbal exchange. In: ETP 2004, pp. 40–44 (2004)

13. Yoshida, N., Aida, D., Fukushima, S., Naemura, T.: Practical study of positive-feedback button for brainstorming with interjection sound effects. In: CHI 2016, pp. 1322–1328 (2016)
14. Yumura, T., Lim, Y., Tan, Y.: PICALA: an interactive presentation system to share reaction of audiences with light color. In: Asian CHI Symposium 2016, pp. 46–52 (2019)
15. YouTube. https://www.youtube.com/. Accessed 01 July 2022
16. Vue.js. https://jp.vuejs.org/index.html. Accessed 01 July 2022
17. Zoom. https://zoom.us/. Accessed 01 July 2022

Support by Visually Impaired: A Proposal for a System to Present Walkability on Maps Using White Cane Data

Rinta Hasegawa(✉) and Junko Ichino

Tokyo City University, 3-3-1, Ushikubonishi, Tsuzuki-ku, Yokohama 224-8551, Kanagawa, Japan
g2293102@tcu.ac.jp

Abstract. Since visually impaired people cannot rely on their sense of sight, they use other senses, such as hearing touch, and smell, more keenly than sighted people. They may also have superior senses and skills to sighted people. However, most previous studies on visually impaired people considered them as people who receive support, such as in transportation, reading, and writings. We consider visually impaired people as "givers" of support rather than "recipients" of support and explore frameworks that utilize the abilities and skills of the visually impaired to support sighted people. We recognize that visually impaired people use a white cane to search for easy-to-walk areas without steps and obstacles. Accordingly, we propose a system that supports sighted people by presenting the walkability of sidewalks on a map based on the white cane operation history of visually impaired people. We first conducted a survey and two preliminary studies, and then derived system requirements based on the findings. Next, we designed a system based on these requirements and confirmed that the system detects some behaviors related to walking difficulty from the history of white cane operations.

Keywords: Visually impaired · White cane · Walkability maps · Support for walking

1 Introduction

According to the data presented by World Blind Union in 2010, 285 million people worldwide have been identified as visually impaired [1]. Visual impairment has two types according to the degree of disability: total blindness, where all visual acuity is lost, and low vision, where a portion of the visual field is lost or vision is blurred. The visually impaired are also divided into two types according to the time of onset: congenitally visually impaired, born with visual impairments, and those with acquired visual impairments by diseases or accidents.

The visually impaired face difficulties in various situations, especially in transportation, reading, and writing. The gap between the visually impaired and the sighted is widening because of the recent trends toward greater emphasis on "visualization." Thus,

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 141–157, 2022.
https://doi.org/10.1007/978-3-031-20218-6_10

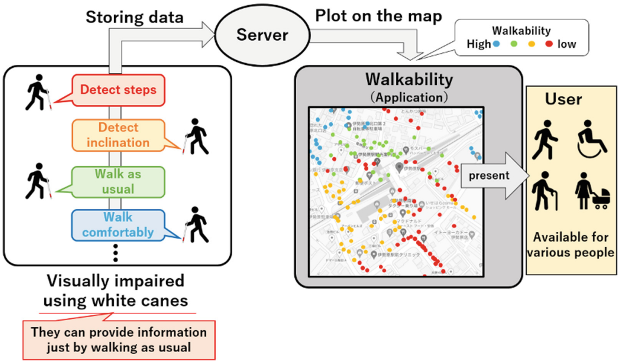

Storing data
Server
Plot on the map
Walkability
High
low
Detect steps
Detect inclination
Walk as usual
Walk comfortably
Visually impaired using white canes
They can provide information just by walking as usual
Walkability (Application)
present
User
Available for various people

Fig. 1. System overview. (Left: Visually impaired perceive obstacles using white cane. Right: Sighted plan their route referring to the walkability map where markers are colored depending on the walkability level.)

many studies have been conducted in support of activities that are essential for the visually impaired in their daily lives, such as transportation [2, 5, 13], reading [3], and writing [4]. Recently several studies have focused on enriching their lives, such as children's play [6] and taking pictures [7]. Most of the previous studies identify the visually impaired as "recipients" of support.

From a different perspective, it could be considered that the visually impaired have superior abilities because of their visual impairment. Because they cannot rely on visual modalities, they effectively use other sensory modalities such as auditory modalities. Congenitally visually impaired are considered to have better hearing than the sighted. In other words, the visually impaired can be regarded as "givers" of support. However, few studies have explored the possibilities of utilizing their superior senses and skills.

We consider the visually impaired as "givers" of support, instead of "recipients" of support, and propose a method to support the sighted using information available only to the visually impaired. We developed a system that detects walking difficulty on roads based on visually impaired data obtained using a white cane and presents the walkability on maps for the sighted, such as the elderly and wheelchair users. Our system targets the visually impaired who are congenitally visually impaired with low dependence on visual information.

The rest of this paper is organized as follows: we describe related works in Sect. 2, preliminary surveys and studies conducted in Sects. 3 through 5, the implemented system in Sects. 6 and 7, and initial testing of the implemented system in Sect. 8.

2 Related Work

2.1 Visually Impaired as Recipients of Support

As mentioned in the Introduction section, various studies have been conducted to support visually impaired.

Transportation Support. Transportation is difficult for the visually impaired who cannot acquire sufficient visual information and have difficulty in understanding their surroundings. Most of them use a white cane to perceive information about their surroundings. However, this method is insufficient. For example, the white cane cannot detect obstacles such as signboards that may hit the upper body when walking. Moreover, other problems exist, such as collision with cars on the roadway or trains at station platforms. To solve these problems, so-called Smart canes [8, 9], which present information using sound and vibration, have been commercialized, but they are not widely used.

In previous studies, many systems that use sound and tactile sensations to provide environmental information have been developed to improve the walking safety of the visually impaired. In the study of Kayukawa et al. [2], a suitcase-type device that produced audible warnings to avoid collisions was developed.

Reading and Writing Support. Reading and writing are also difficult tasks for visually impaired because they cannot acquire visual information. The inability to recognize signs indicating destinations in locations such as train stations and roads is a major factor that make transportation difficult for the visually impaired.

Shilkrot et al. [3] developed a system to allow visually impaired to read by following correctly the text using a finger extension. Feiz et al. [4] developed a system to help visually impaired to write letters by themselves. The system uses speech to present the instructions for filling in the correct position in the document.

Studies on Other Support Types. In addition to transportation, reading, and writing, many other support types have been developed recently to enrich the life of visually impaired.

Abreu et al. [6] developed a play system using information technology that helps visually impaired children learn logical thinking and algorithms, which was constructed as a playful game. Oshima et al. [10] developed a system that enabled visually impaired spectators to grasp the game situation of a blind soccer match by presenting the situation using a Braille display. Numerous other systems have been developed to support the actions of visually impaired, such as support for taking pictures using a camera [7], music creation [11], and playing a game [12].

2.2 Visually Impaired as Givers of Support

While there are numerous studies regarding visually impaired, most aim only to support them because they are often perceived requiring support.

However, they have unique skills that are not available to sighted. Ito [14] elaborated on the way visually impaired see the world using episodes obtained from actual visually impaired, which stated that it is possible for visually impaired to support sighted using their unique skills.

In an event called “Social View” [14, 15], art works were interactively viewed by the visually impaired and sighted, in which multiple people exchanged their opinions during the exhibition. The purpose of this event was not merely to provide information to the visually impaired about the artwork through explanations given by a sighted person. By having visually impaired participate in the interactive viewing, sighted participants were asked to verbalize things that are not normally required. Moreover, through responding to simple questions presented by the visually impaired, sighted were able to update their own views and perceptions of the artworks. That is, at this event, visually impaired participants provided new insights about art works to sighted participants.

"Dialogue in the Dark" [16] is an event in which visually impaired are in a position to support the sighted. This event is implemented in the dark space where the visually impaired act attendants and the sighted act as participants. Owing to being deprived of visual information, sighted participants can gain new insights that they would not normally be able to obtain.

Thus, while previous studies have considered the visually impaired as requiring support, it is possible for visually impaired to support sighted using their unique skills. Although there have been several such events, there have been few studies on systems that use information possessed by the visually impaired to assist the sighted.

3 Survey on Attitudes of the Visually Impaired as Givers of Support

To develop a system supporting sighted, it is necessary to clarify whether the visually impaired would like to be in a position to support the sighted and whether the visually impaired have had any experience of providing a unique insight to sighted. To clarify these issues from the visually impaired viewpoint, we conducted an interview survey with visually impaired working at a blind school.

3.1 Method

The survey was conducted by telephone in May–June 2021 for about 20 min each. Table 1 presents the data regarding the interviewees who participated in the survey, while Table 2 lists the main questions. This survey consisted of a basic questionnaire regarding the tools often used by the visually impaired and problems they face. Considering the possibility of offending the participants, we did not ask about their degree of visual impairment and recorded them only after disclosure during the conversation.

Table 1. Interviewees who participated in the survey.

ID	Gender	Degree of visually impairment	ID	Gender	Degree of visually impairment
i1	Female	Acquired blind	i5	Male	Low vision
i2	Male	Congenital blind	i6	Male	Low vision
i3	Male	Blind	i7	Male	Not disclosed
i4	Male	Low vision			

Table 2. Main questions and answers.

No	Questions and answers
Q1	**Would you like to be in a position to assist sighted?**
A1	Positive answers (6/7) "I would be happy if I could be of help (i1)" "I should have a relationship with them (i3)" etc.
Q2	**What kind of information do you usually use to get around?**
A2	Unique ways of using information (4/7) "I use the characteristic smell of a bakery as a landmark (i1)" "Walking parallel to the sound of cars on the road (i2)" etc.
Q3	**Do you have confidence in information other than visual information?**
A3	Positive responses (3/7) "Confidence in subtle sounds (i1)" "I have good attention to sound (i4)" etc.
Q4	**Have you ever been taught or made aware of information that was not available to the sighted?**
A4	Applicable experiences (6/7) "I noticed differences in some steps on a staircase that sighted had not noticed(i1)" "I noticed the sense of season from the smell of the air (i2)" etc.

3.2 Results and Consideration

Table 2 lists the main questions and answers of this survey.

In this survey, several questions were asked to clarify whether the visually impaired can support the sighted. This survey revealed that the visually impaired would like to be in a position to assist the sighted and the blind and visually impaired mainly use information gathered with their own unique skills. In addition, experiences of assisting the sighted, such as "Social View" and "Dialogue in the Dark" described in Sect. 2, were experienced by many visually impaired. That is, it is possible for the visually impaired to support sighted.

4 Preliminary Study on Awareness of Visually Impaired

For system implementation, it is necessary to clarify whether visually impaired actually notice what are not noticed by sighted and, if so, what kind of information they notice. Accordingly, we conducted a preliminary study to observe and compare methods of acquiring information employed by the visually impaired and sighted to clarify whether the visually impaired perceive information more advantageously than the sighted in certain scenarios. In addition, a study is conducted to analyze whether there are characteristic behaviors regarding information search and awareness of the visually impaired to design the system requirements.

4.1 Method

There were five participants: two sighted and three visually impaired (two blind and one low vision) (Table 3). The two visually impaired participants were asked to participate in the study using a white cane as usual.

Figure 2 depicts the study situation. A traditional Japanese garden in a park (Fig. 3), was selected as the study site because it was relatively easy to walk around and was rich with information, including visual information and the sounds of creatures and water. The route (red line in Fig. 3) would take the participants several minutes by normal walking. To ensure that both sighted and visually impaired participants could walk along, the study was conducted assisted by the staff. Video and audio (EEG data using an EEG sensor only for P2 [17]) were recorded. In addition, as an study task, participants were asked to verbally report everything they noticed during the walk, the content of which was also recorded. The study was approved by the Ethical Review Committee for Research on Human Subjects of Tokyo City University (Approval No.: 2021-h07).

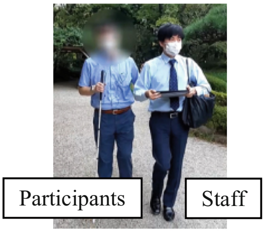


Fig. 2. Study setting.

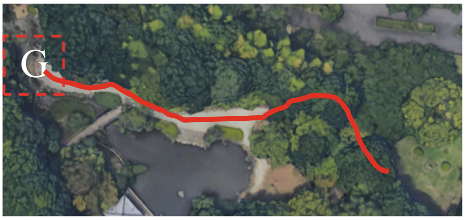


Fig. 3. Study location and route.

4.2 Results and Consideration

Awareness. Table 3 presents the number and content of the findings reported by the participants. The results demonstrate that visually impaired participants obtained information that sighted participants did not notice.

Behavior. The visually impaired participant, who is totally blind, was observed to intentionally move his white cane widely to search for information, even though he was less dependent on the white cane owing to the presence of his companion. It is thought that observing the white cane movement of a person with visual impairment who walks alone in a detailed manner may enable us to confirm movement synchronization to that of a visually impaired.

Biological Response. We analyzed the data collected by the EEG sensor attached to P2 to assess whether there were any changes in the values associated with awareness; however, we identified almost no characteristic data.

5 Preliminary Study on Walking Behavior of Visually Impaired Using a White Cane

From the preliminary study on awareness of visually impaired, we determined that it was possible to extract behaviors related to environmental awareness from the movements of the visually impaired using a white cane. Therefore, we observed the visually impaired walking with a white cane to extract behaviors associated with their environmental awareness.

Table 3. Results of study tasks for visually impaired and sighted participants.

ID	Gender	Visually impaired or sighted	Date	Time	Number and content of awareness (Categorized by sense [18])
P1	Female	Visually impaired (acquired blind)	2021/8/11	3 min 47 s	**20 pieces** "Sound of water" "Pebbles below" etc. (Auditory 50%, skin sensation 30%, multiple sensations 20%)
P2	Male	Visually impaired (congenital blind)	2021/9/14	3 min 47 s	**16 pieces** "Sound of water" "In a narrow space" etc. (Auditory 37.5%, equilibrium 12.5%, skin sensation 18.8%, multiple senses 31.3%)
P3	Male	Visually impaired (low vision)	2021/8/11	2 min 12 s	**7 pieces** "Spread" "wind" etc. (Auditory 14.3%, equilibrium 14.3%, skin sensation 57.1%, multiple senses 14.3%)
P4	Male	Sighted	2021/7/21	2 min 3 s	**16 pieces** "Green curtains" etc. (Visual 75%, equilibrium 6.3%, multiple senses 18.6%)
P5	Male	Sighted	2021/7/11	2 min 30 s	**15 pieces** "Lots of trees" "Clean water" etc. (Visual 86.6%, auditory 6.6%, multiple senses 6.6%)

5.1 Method

P1 and P2 were blind visually impaired who used white canes in the preliminary study. They were asked to walk with a white cane on a straight sidewalk near their workplaces. The participants were videotaped while walking, using modified white canes incorporated with a microphone and acceleration sensor [20] (Fig. 4). The accelerometer was attached to extract characteristic data of walking with a white cane for later analysis, and the microphone was attached to match the time information of the acceleration data. The study was approved by the Ethical Review Committee for Research on Human Subjects of Tokyo City University (Approval No.: 2021-h07).

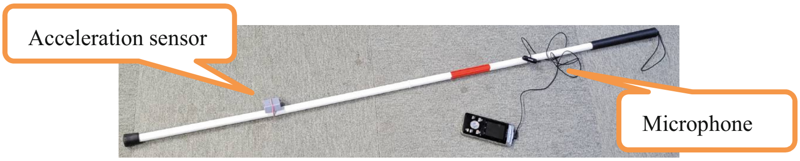


Fig. 4. Modified white cane with sensors.

5.2 Results and Consideration

Two participants performed periodic movements with the white cane, varying their walking speeds and occasionally stopping. The periodic movements performed by the participants were primarily a technique called "touch technique" and partially a technique called "slide technique" [19]. These techniques were used to ensure that there was enough space for them to pass. The touch technique that the participant primarily used was the technique of alternating the white cane on the left and right side of the front in accordance with the gait tempo. On the other hand, the slide technique involved sliding forward from side to side while keeping the white cane on the ground. Table 4 lists the characteristic movements observed in P1 and P2 during the preliminary study, along with the factors that were assumed to have caused these movements.

The two locations (1–2 and 1–3) where characteristic movements were observed in walking patterns of P1 were the same for P2 (2–2 and 2–4), respectively. Figure 5 shows P1 walking at location 1–2, and Fig. 6 shows P2 walking at location 2–4.

Table 4. Characteristic behavior of the participants and presumed factors.

ID	No	Characteristic behavior	Presumed factors
P1	1–1	Changing the touch techniques for the white cane	Unknown
	1–2	Changing the touch techniques for the white cane	The sidewalk is leaning toward the roadway at the entrance/exit of the parking lot Texture is changing
	1–3	Changing the touch techniques for the white cane	Protuberance of the sidewalk due to street trees and root growth
	1–4	Changing the touch techniques for the white cane	Protuberance of the sidewalk due to street trees and root growth
P2	2–1	Changing the touch techniques for the white cane	Unknown
	2–2	Deceleration and intermittent stopping	The sidewalk is leaning toward the roadway at the entrance/exit of the parking lot Texture is changing
	2–3	Deceleration of walking	The sidewalk is leaning toward the roadway at the entrance/exit of the parking lot
	2–4	Deceleration of walking	Protuberance of the sidewalk due to street trees and root growth
	2–5	Deceleration and intermittent stopping	Unknown
	2–6	Stop walking	Sound of cars passing nearby

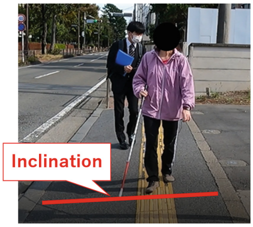


Fig. 5. P1 perceiving inclination of the road and adjusting the white cane movement.

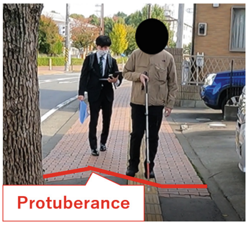


Fig. 6. P2 perceiving protuberance of the road and adjusting the white cane movement.

6 System Requirements

The following section summarize the findings of the preliminary studies on awareness and walking behavior of the visually impaired. In fact, there is an abundance of non-visual information that can be obtained only by visually impaired (Sect. 4). The white cane movement patterns of visually impaired change from periodic to characteristic movements, such as changes in walking speed, intermittent stopping, and changes in the white cane movements (Sect. 5).

Accordingly, we determined that the information regarding the walking difficulty, which can be acquired only by visually impaired, is expressed in the walking behavior using a white cane. This study proposes a system that presents walkability on maps by utilizing data regarding the use of a white walking cane (Fig. 1). The following are the system requirements:

I. Detect walking motions related to walking difficulty from the data regarding the use of a white cane (system input part).
II. The system presents information regarding walking difficulty on a map for sighted (system output part).

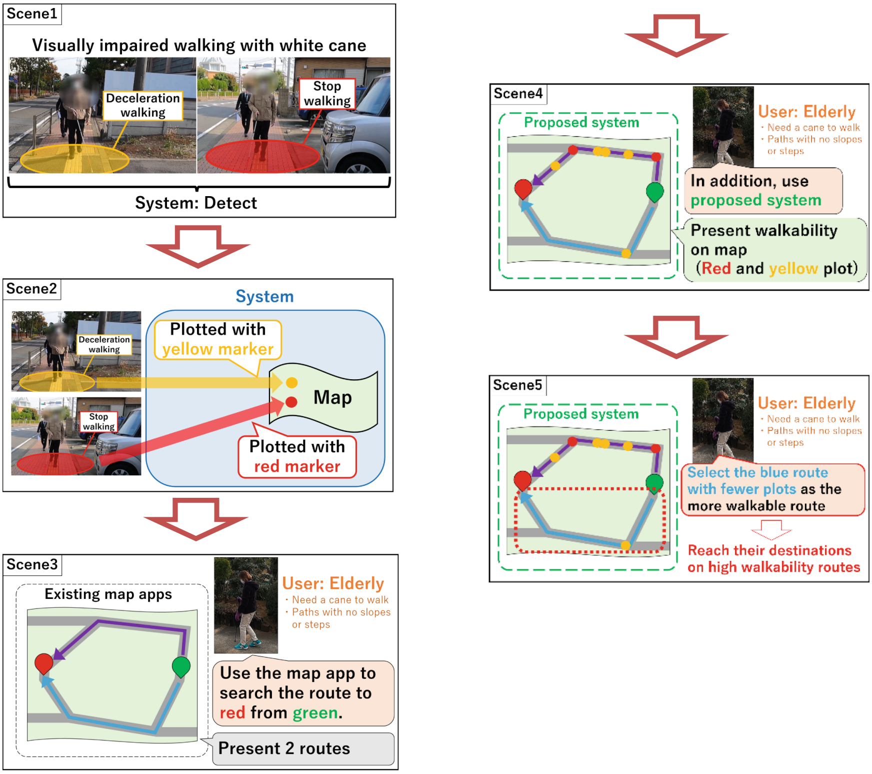

Scene1
Visually impaired walking with white cane
Deceleration walking
Stop walking
System: Detect
Scene2
System
Deceleration walking
Stop walking
Plotted with yellow marker
Map
Plotted with red marker
Scene3
Existing map apps
User: Elderly
・Need a cane to walk
・Paths with no slopes or steps
Use the map app to search the route to red from green.
Present 2 routes
Scene4
Proposed system
User: Elderly
・Need a cane to walk
・Paths with no slopes or steps
In addition, use proposed system
Present walkability on map (Red and yellow plot)
Scene5
Proposed system
User: Elderly
・Need a cane to walk
・Paths with no slopes or steps
Select the blue route with fewer plots as the more walkable route
Reach their destinations on high walkability routes

Fig. 7. Example usage scenario.

Figure 7 presents an example usage scenario for the system. This scenario assumes that the system is used by an elderly person who needs a cane for walking. In the figure, the elderly user uses the system to plan an easy-to-walk route from the green marker to the red marker. It is also assumed that many visually impaired users provide data to the system and sufficient information on walkability is displayed on the map.

7 System Design and Implementation

The proposed system is presented in Fig. 1, and a diagram of the entire system is shown in Fig. 8.

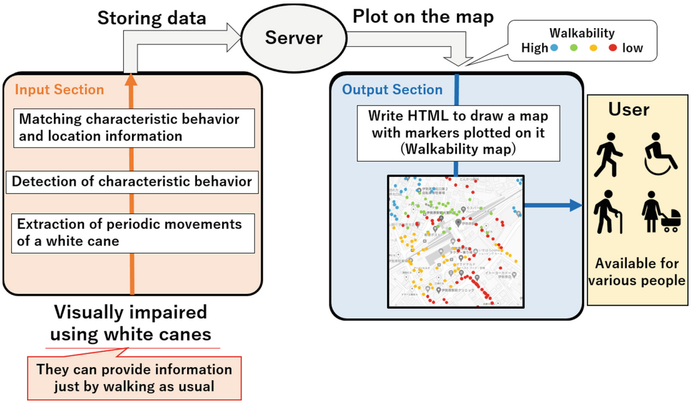

Storing data
Server
Plot on the map
Walkability
High ● ● ● ● low
Input Section
Matching characteristic behavior and location information
Detection of characteristic behavior
Extraction of periodic movements of a white cane
Output Section
Write HTML to draw a map with markers plotted on it (Walkability map)
User
Available for various people
Visually impaired using white canes
They can provide information just by walking as usual

Fig. 8. System diagram.

7.1 Design of System Input

To design the input system that satisfies System Requirement I, we extracted walking behaviors related to the difficulty experienced by visually impaired that the system should detect based on the observation study (Sect. 5) regarding walking behavior using a white cane (Table 5). For example, “a visually impaired stopped walking because he/she felt danger” is extracted from the study results when the characteristic walking behavior occurred. In this system, the visually impaired can become information providers simply by attaching a sensor to their white cane and proceeding with their daily lives as usual.

7.2 Implementation of System Input

A small accelerometer is attached to a white walking cane to monitor the walking motions related to the difficulty experienced by visually impaired, as presented in Table 5. The data related to walking difficulty can be collected when visually impaired use the white walking cane as usual. A GPS receiver GR-7BN is used to acquire location information, which is connected to a PC via a USB. Subsequently, a dedicated application outputs position and time information in CSV format at 1/10 s intervals.

The system detects the most basic walking motion patterns reported in Table 5: the touch technique cycle (A in Table 5), stopping (E in Table 5), and change in speed (D in Table 5). The cycle of the touch technique is recorded as the white cane repeatedly touches the ground. Figure 9 shows a diagram of the implemented input unit.

Table 5. Movements of visually impaired related to walking difficulty.

	Movements in walking
A	Thrusting the ground with the white cane (touch technique)
B	Swinging the white cane to the left
C	Swinging the white cane to the right
D	Changing the walking speed
E	Intermittent stopping
F	Contact between white cane and step
G	Contact between white cane and wall
H	Dragging the white cane

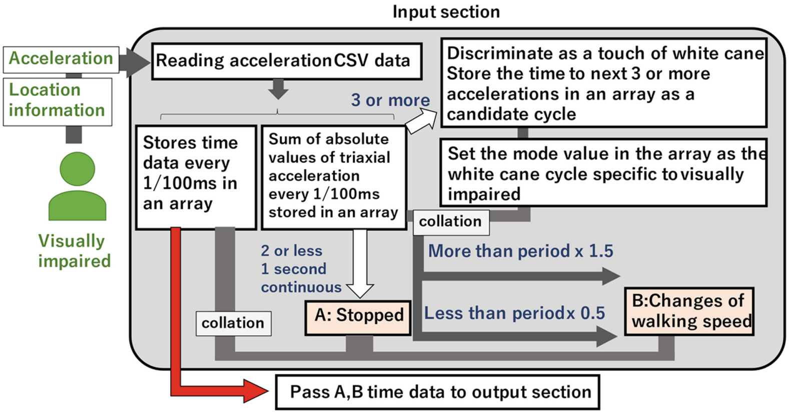

Input section
Acceleration
Location information
Visually impaired
Reading acceleration CSV data
3 or more
Discriminate as a touch of white cane Store the time to next 3 or more accelerations in an array as a candidate cycle
Stores time data every 1/100ms in an array
Sum of absolute values of triaxial acceleration every 1/100ms stored in an array
Set the mode value in the array as the white cane cycle specific to visually impaired
collation
More than period x 1.5
2 or less 1 second continuous
Less than periodx 0.5
B:Changes of walking speed
collation
A: Stopped
Pass A,B time data to output section

Fig. 9. Input system diagram.

7.3 Design of System Output

In this section, a system that presents the paths and routes that are easy to walk is implemented by displaying four marker levels that are presented in Table 6 on a map, so that the output system satisfies System Requirement II.

KCS Corporation's "Creating 'Walkability Maps' for Policy Evaluation of Walking Space and Community Development" has attempted to create a walkability map in a similar manner [21]. This initiative aims to create a walkability map to improve the walkability in city a location and understand the problems in the walking space. The map is color-coded according to each stage. A similar study that is conducted by Prandi et al. on accessible navigation apps [22].

The markers presented by this system, along with map information, are intended for two main uses. First, to plan routes to destinations. This allows the user to use the map while walking along the generated route. The second is to check for traces of characteristic behavior of the visually impaired in the vicinity.

Table 6. System output marker type.

Walkability	Color
Very easy to walk	Blue
Easy to walk	Green
Hard to walk	Yellow
Very hard to walk	Red

7.4 Implementation of System Output

To satisfy system requirements, the Google Maps API is used. This is a service that allows to customize the functions provided by Google Maps and embed them in websites. Using this service, the location data regarding characteristic movements are visually presented by displaying the markers reported in Table 6. Figure 10 shows a output system diagram. Figure 11 shows a map presented by the system using the acceleration data of participant P2's white cane collected while his walking.

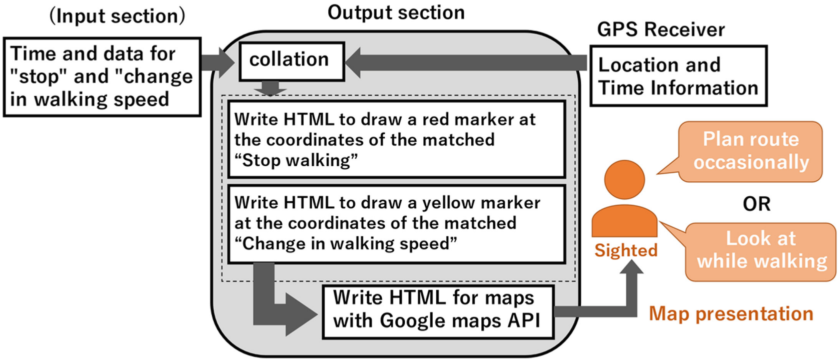

(Input section)
Time and data for "stop" and "change in walking speed
Output section
collation
Write HTML to draw a red marker at the coordinates of the matched "Stop walking"
Write HTML to draw a yellow marker at the coordinates of the matched "Change in walking speed"
Write HTML for maps with Google maps API
GPS Receiver
Location and Time Information
Plan route occasionally
OR
Look at while walking
Sighted
Map presentation

Fig. 10. Output system diagram.

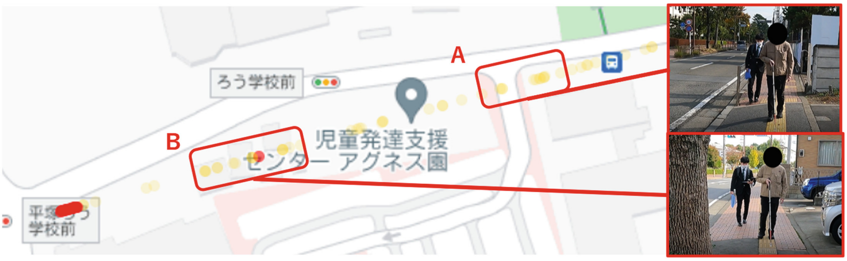


Fig. 11. Walkability map presented using P2 data.

8 Determining Behavior of Visually Impaired from White Cane Sensor Data

8.1 Initial Testing Using White Cane Usage Data Obtained from Preliminary Study on Walking Behavior (Sect. 5)

In the implemented system, the acceleration data are visualized in a graph to determine the moment when the white cane is thrust. Then, when the acceleration becomes significantly large is set as the point where the white cane touches the ground. It is necessary to verify whether this value setting can correctly detect the motion of thrusting a white cane.

We compared the videos of two participants recorded during the preliminary study on walking behavior (Sect. 5) with the system output using the collected acceleration data related to the two participants. Then, we measured and compared the number of times they thrust the white walking cane in the videos and number of times the system detected the thrust. The comparison results are reported in Table 7, which demonstrate that the system was able to detect the white cane motion for P1 and P2 with high accuracy values of 93% and 100%, respectively. P2 is a visually impaired participant who is congenitally blind, whose white cane operation method is sophisticated and blur-free; thus, the system can detect his movements accurately.

Table 7. Cane touch detection accuracy.

ID	Actual counts	Counts by system	Accuracy
P1	191	178	**93%**
P2	342	342	**100%**

8.2 Initial Testing Using New White Cane Usage Data

As mentioned previously, the implemented system was able to detect white cane movements with high accuracy in the dataset constructed with data observed from the participants. However, individual differences must be considered also. Accordingly, two new participants, namely, P1 and P2 (Table 8) who are blind, visually impaired, and use a white cane, were employed to verify whether a similar dataset can be observed and displayed on the map with new participant when they traversed along the route considered in the preliminary study on walking behavior (Sect. 5).

Output part of the system using the participant data collected in this testing is shown in Fig. 12 and 13. The system was able to detect a periodic motion even when new participants were observed, suggesting that the system can function despite individual differences. Similar to the results obtained with the participants in the preliminary study on walking behavior (Sect. 5), periodic movements near the entrance and exit of the parking lot (Fig. 5) were detected with high accuracy for both P1 and P2. This indicates that this is a place where many visually impaired feel difficulties in walking.

Table 8. Participants of the initial testing using new white cane usage data.

ID	Gender	Degree of visual impairment	Date
p1	Male	Blind	2021/1/19
p2	Male	Blind	2021/1/20

In addition, two participants used not only the touch technique but also the sliding technique in the early study phase, during which they kept the white cane touching the ground and dragged it back and forth from side to side. The two participants in the preliminary study on walking behavior (Sect. 5) always used the touch technique. Hence, the implementation of the system was mainly based on the extraction of this touch technique cycle. This may have caused classifying the movement patterns exhibited by P1 and P2 in the early phases, namely, the slide technique, as periodic movements.

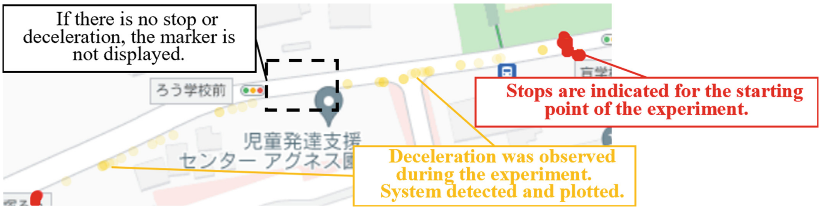

If there is no stop or deceleration, the marker is not displayed.
ろう学校前
児童発達支援センター アグネス園
Stops are indicated for the starting point of the experiment.
Deceleration was observed during the experiment. System detected and plotted.

Fig. 12. Walkability map presented using p1 data.

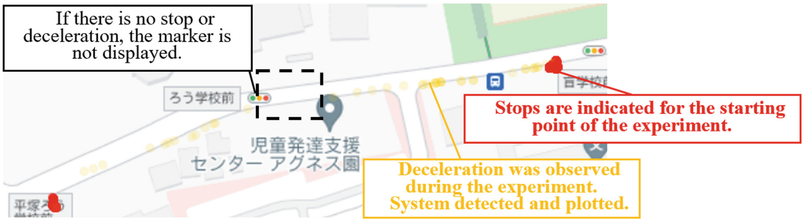

If there is no stop or deceleration, the marker is not displayed.
ろう学校前
児童発達支援センター アグネス園
Stops are indicated for the starting point of the experiment.
Deceleration was observed during the experiment. System detected and plotted.

Fig. 13. Walkability map presented using p2 data.

9 Conclusion and Future Work

After conducting a survey and two preliminary studies, we implemented a system that produces a walkability map to aid the sighted using data from white cane usage patterns exhibited by the visually impaired. This system creates a scenario where the visually impaired can support the sighted, which has hardly been researched in previous studies.

In the survey and preliminary studies conducted before system implementation, we determined that the visually impaired can perceive non-visual information better than the sighted. They perceive information based on how they use the white cane, which is difficult for the sighted.

Initial testing demonstrated that the system could extract individual periodic and detect non-periodic movements despite individual differences among the visually impaired. However, this system might not be compatible with the sliding technique, a white cane movement technique used by the visually impaired.

The implemented system is still in its first version, and there is significant room for improvement. First, the current system can only detect the temporal cycle of the touch technique of a white cane, and it operates only with the absolute sum of triaxial acceleration and time information. This implies that only three of the eight items in Table 5 can be classified, and we must add a function to classify the remaining five items in Table 5, for example, to discriminate sliding techniques. Currently, we are investigating the use of support vector machines for classification based on the videos and acceleration data of the white cane collected from the participants in the preliminary study on walking behavior (Sect. 5). The classification matches the video and acceleration data and uses the triaxial acceleration data of the white cane swung in each direction as training data. The classification adds not only left-right motion discrimination but also the rest of the classification functions listed in Table 5, and we plan to start work on this in the future.

Second, this study assumes that the system's walkability maps will be used by many visually impaired people, including those with walking difficulties, wheelchair users, and the elderly. Therefore, it is necessary to investigate and analyze the walkability information that various system users would like to know. Currently, interviews are being conducted with a community where various elderly people belong, and their demands regarding the walkability of the paths are being investigated.

Third, after improving the system and making it practical, we must evaluate the walkability map presented, by collecting usage data from many visually impaired users.

Acknowledgment. We would like to thank the staff of a school for the blind in Kanagawa Prefecture, students of Tokyo City University, and others for their cooperation in the surveys and studies.

References

1. World Blind Union: Blindness Global Fact Sheet, October 2011 (2011)
2. Kayukawa, S.: BBeep: a sonic collision avoidance system for blind travelers and nearby pedestrians. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI 2019), no. 52, pp. 1–12. ACM (2019)
3. Shilkrot, R., Huber, J., Meng Ee, W., Maes, P., Nanayakkara, S.: FingerReader: a wearable device to explore printed text on the Go. In: Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems (CHI 2015), pp. 2363–2372. ACM (2015)
4. Feiz, S., Masum Billah, S., Ashok, V., Shilkrot, R., Ramakrishnan, I.: Towards enabling blind people to independently write on print forms. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI 2019), no. 300, pp. 1–12. ACM (2019)
5. Alkhanfer, A.A., Ludi, S.: Visually impaired orientation techniques in unfamiliar indoor environments: a user study. In: Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2020), pp. 283–284. ACM (2020)

6. Abreu, L., Cristina Pires, A., Guerreiro, T.: TACTOPI: a playful approach to promote computational thinking for visually impaired children. In: Proceedings of the 16th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2014), no. 82, pp. 1–3. ACM (2014)
7. Vazquez, M., Steinfeld, A.: Helping visually impaired users properly aim a camera. In: Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2012), pp. 95–102. ACM (2012)
8. WeWALK. https://wewalk.io/en/. Accessed 7 July 2022
9. Amedia Smart cane SC1. https://www.amedia.co.jp/product/walking/cane/white-cane/SC1.html. Accessed 7 July 2022
10. Ohshima, H., Kobayashi, M., Shimada, S.: Development of blind football play-by-play system for visually impaired spectators: tangible sports. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI 2021), no. 209, pp. 1–6. ACM (2021)
11. Payne, C.W., Yixuan Xu, A., Ahmed, F., Ye, L., Hurst, A.: How blind and visually impaired composers, producers, and songwriters leverage and adapt music technology. In: Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2020), no. 35, pp. 1–12. ACM (2020)
12. Walia, A., Goel, P., Kairon, V., Jain, M.: HapTech: exploring haptics in gaming for visually impaired. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI 2020), pp. 1–6. ACM (2020)
13. Moreno, M., Shahrabadi, S., Jose, J., du Buf, J.M.H., Rodrigues, J.M.H.: Realtime local navigation for the blind: detection of lateral doors and sound interface. In: Proceedings of the 4th International Conference on Software Development for Enhancing Accessibility and Fighting Info-exclusion (DSAI 2012), vol. 14, pp. 74–82. Science Direct (2012)
14. Ito, A.: How do blind people see the world?, 1st edn. Kobunsha Shinsyo (2015)
15. Anzai, Y., Hirano, T., Yamada, S., Shiose, T.: Art appreciation through discussion with visually impaired. J. Assoc. Art Educ. (39), 27–38 (2018)
16. Dialogue in the Dark. https://did.dialogue.or.jp/. Accessed 12 Jan 2022
17. NeuroSky mindwave mobile 2. https://www.neurosky.jp/mindwave-mobile2/. Accessed 14 Jan 2022
18. Katano, Y.: Shinteiban Zukai One Point Seirigaku. Saio syuppan (2017)
19. Yamada,Y.: Hakujou Hokou Support Hand Book. Dokusyo koubo (2010)
20. Monowireless TWELITE2525A. https://mono-wireless.com/jp/products/TWE-Lite-2525A/index.html/. Accessed 14 Jan 2022
21. Ltd. KCS. http://www.kcsweb.co.jp/works/socialpolicy/. Accessed 16 Jan 2022
22. Prandi, C., Barricelli, B.R., Mirri, S., et al.: Accessible wayfinding and navigation: a systematic mapping study. In: Universal Access in the Information Society (2021)

How Teacher Education Students Collaborate When Solving an Asymmetric Digital Task

Meeli Rannastu-Avalos(✉), Mario Mäeots, and Leo A. Siiman

University of Tartu, Ülikooli 18, 50090 Tartu, Estonia
{meeli.rannastu-avalos,mario.maeots,leo.siiman}@ut.ee

Abstract. Collaboration skills are essential when people work together to address challenges in society that are too complex for any one person to solve alone. University teacher education programs should promote teaching practices and tasks that support the development of collaboration skills. In this study, 292 university teacher education students worked in pairs on an asymmetric digital task which required collaboration via chat message exchanges. An asymmetric computer simulation was used to unevenly distribute functionality necessary for solving the task among the pairs, thereby establishing a condition of interdependence. Log data from the simulation and chat messages were collected. However, simple learning analytics indicators such as time spent on the task, the number of questions asked, or the length of messages exchanged did not reveal any strong correlation with task performance. Consequently, qualitative analysis was conducted to review the chat message exchanges and determine what characterises collaborative performance. According to the qualitative findings, sharing information and developing a common understanding were critical for success. We distinguished three unique groups based on coding collaborative behaviour: low, medium, and high performers. Dyads tend to perform poorly when a student does not share their unique information. The asymmetric simulation used in this study is an excellent opportunity to practice developing collaboration skills. Although this was a new problem-solving task for most students, half of the dyads were successful at it. Future research should investigate whether a second attempt with a similar asymmetric digital task would enable all participants to be successful.

Keywords: Collaboration · Interdependence · Teacher education · Asymmetric simulations · Learning analytics · Shared understanding

1 Introduction

Collaboration skills are essential for solving complex problems confronting today's society. But how to best teach collaboration skills? University teacher education programs have sometimes been criticised for presenting theoretical knowledge without demonstrating how such knowledge can be applied in classroom practice to support student learning [1]. A critical competence or general skill promoted by many education stakeholders to guide 21st-century educational policy and practice is that of collaboration. Teacher education programs have a responsibility not only to emphasise the importance

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 158–174, 2022.
https://doi.org/10.1007/978-3-031-20218-6_11

of collaboration in education but to offer teaching practices, example activities, and feedback for developing students' collaboration skills [2].

A large-scale study of students' collaborative problem-solving skills was conducted in 2015 by the Programme for International Student Assessment (PISA) [3], with over 500,000 students from 52 countries participating. The PISA study found that only 8% performed at Level 4 (i.e., the highest performance level). In order to assess collaboration, PISA developed computer-based tasks. These digital tasks were based on a theoretical framework of collaboration skills that included establishing and maintaining shared understanding (e.g., discovering unique perspectives and knowledge of team members), taking action to solve the problem (e.g., identifying and describing subtasks) and keeping the team organised (e.g., understanding roles and monitoring member contributions). However, the PISA digital tasks did not involve human-to-human collaboration. Instead, a student interacted with an adaptive computer algorithm which presented pre-defined chat messages in the context of a virtual environment and from which a student had to select the most appropriate message.

A more authentic approach (i.e., human-to-human collaboration) towards using technology to assess students' collaborative problem-solving skills was applied by the Assessment and Teaching of 21st Century Skills (ATC21S) project [4]. ATC21S developed digital tasks in which two students could partially control different aspects of a computer simulation. The pair of students worked on separate computers but could communicate via a chat messenger app. In one task, a beam balance simulation was displayed to both students, but each student could place masses on only one side of the balance. They therefore had to work together in order to explore where to place masses to obtain a balanced state. This type of digital task introduces asymmetry into the task and is useful in establishing a condition of interdependence between the two collaborators. Johnson and Johnson argue that *positive interdependence* is one of the essential elements of cooperative learning and define it to be "the perception that we are linked with others in a way so that we cannot succeed unless they do" [5].

A few asymmetric simulations are freely available on the internet, for example, on the Go-Lab learning platform [6], and have been developed for use with smartphones in the context of collaborative scientific inquiry [7]. In a previous study using asymmetric simulations with school pupils, it was found that working on asymmetric collaborative inquiry tasks is challenging. Consequently, more systematic instruction and feedback about collaboration skills is needed at schools [8].

Collaboration is a coordinated, synchronous action that occurs due to continuous attempts to create and sustain a shared understanding of a situation [9]. Shared regulation is defined as the creation and construction of joint meaning for work and the negotiation and exchange of ideas about how and in what manner mutual goals for the job might be realised [10].

Collaboration requires communication because it allows learners to develop a mutual understanding of the shared problem space and engage in knowledge production [9, 11]. Indicators of successful collaboration within learner dialogue (speech or chat) have often been the quantity (e.g., number and length of utterances and talk time), as well as heterogeneity and transactivity (e.g., turn-taking and building on each other's reasoning) [12–15]. Learners' interaction logs in technology-enhanced learning environments, along with their dialogue, are frequently used in collaborative learning to pinpoint productive actions and patterns [14, 16–20].

2 Aim and Research Questions

Collaboration takes place when a continuous effort is made to establish and maintain a shared understanding of a problem. This is done through systematic collaboration. Socially shared regulation refers to groups managing metacognition together through negotiating, iterative adjustments to the cognitive, behavioural, motivational and emotional conditions [21]. There are two forms of shared regulation. The first one involves the creation and construction of joint meaning. This involves negotiation and exchange of ideas as to how and in what manner the mutual goals of the assignment might be achieved.

In order for learners to collaborate effectively, there needs to be clear communication between them. This is necessary for developing a mutual understanding of the shared problem and to be able to engage in knowledge production at the same time. Researchers have found that communication processes recorded in chat appear to include indicators of how successful collaboration happens [16, 17, 19, 20]. Therefore, in order to investigate how previous research on collaboration relates to a digital task involving an asymmetric simulation, the following research questions were formulated:

RQ 1. How are simple learning analytics indicators related to collaborative performance on an asymmetric digital task?
RQ 2. How does qualitative content analysis of the collaborative process explain performance on the asymmetric digital task?

3 Method

3.1 Participants

This analysis focuses on 12 teacher education groups taught by the two university teachers at the University of Tartu, Estonia. The groups were pre-service teacher education classes. In total, there were 292 students and 146 dyads. The years of data collection ranged from 2020 to 2022.

3.2 Materials

The Go-Lab Ecosystem, https://graasp.eu, learning platform was used to conduct the lesson and provide an environment for collaboration. The Collaborative Seesaw Lab ("seesaw lab" or "seesaw activity") is an online lab for two students to work collaboratively at a distance (see Fig. 1). The lab has two versions (A and B) in which each student has access to one side of the seesaw. There is a fixed number of objects to put on the seesaw and the lab aims to create a collaborative experience between two students. The partner's side of the seesaw is hidden; thus, the student cannot control it or guess the moves made on the other side. Students access the simulation via a provided URL link or QR code with their mobile device and then enter a room number. The dyads and versions A and B have been assigned before the class. With a common room number, a pair of students enter a joint simulation and chat.

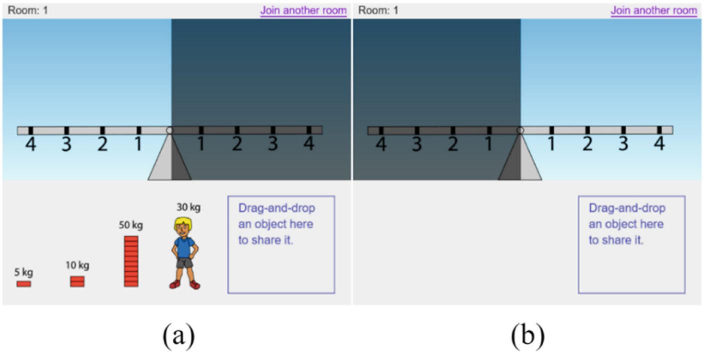
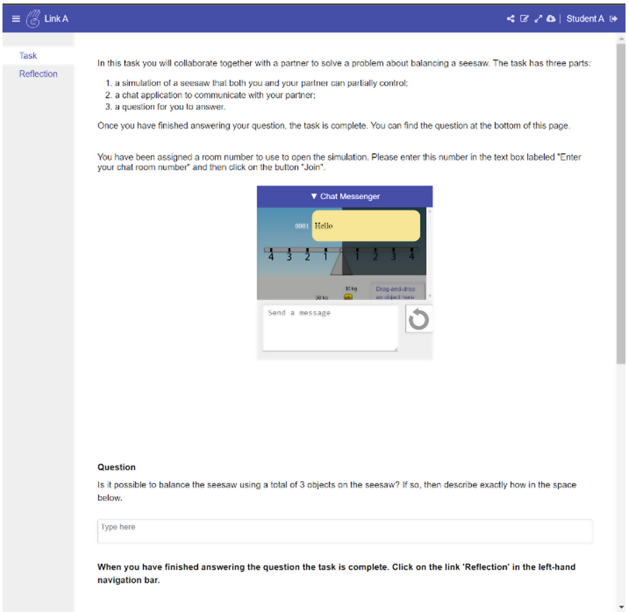


Fig. 1. Screenshot of the Collaborative Seesaw Lab (https://www.golabz.eu/lab/seesaw-lab) used in this study: (a) Version A of the simulation initially contains four masses that can be placed only the left-hand side of the seesaw, (b) Version B of the simulation does not initially contain any masses, but once Student A shares a mass, then it can be placed only the right-hand side of the seesaw.

Students had to complete a problem-solving task; based on the logic of asymmetric collaboration, the dyads share the same simulation that divides into two parts and the problem must be solved together (see Fig. 2). Dyads must find a solution to balance the seesaw with two *and* three objects. At the start, dyads have objects only on the student A-side, and Student A must share them using a green box in the simulation (see Fig. 1).



In this task you will collaborate together with a partner to solve a problem about balancing a seesaw. The task has three parts:

1. a simulation of a seesaw that both you and your partner can partially control;
2. a chat application to communicate with your partner;
3. a question for you to answer.

Once you have finished answering your question, the task is complete. You can find the question at the bottom of this page.

You have been assigned a room number to use to open the simulation. Please enter this number in the text box labeled "Enter your chat room number" and then click on the button "Join".

0001 Hello

Question

Is it possible to balance the seesaw using a total of 3 objects on the seesaw? If so, then describe exactly how in the space below.

When you have finished answering the question the task is complete. Click on the link 'Reflection' in the left-hand navigation bar.

Fig. 2. Screenshot of the asymmetric digital task as seen by Student A in the Graasp (https://graasp.eu) learning environment. Here a chat messenger app has been embedded into the simulation so that collaboration can occur via chat messaging. Student B sees almost the same task in Graasp, except that the simulation is the version B variant, and the question is slightly different "Is it possible to balance the seesaw using a total of 2 objects on the seesaw?".

3.3 Procedure

Since Spring 2021, learning was online at our university, and accordingly there were two types of setting in our study: classes that took place in Spring 2020 and 2022 were held face-to-face (95 dyads participated in this situation) and classes in Spring 2021 were done using synchronous Zoom video lessons (51 dyads were in this situation).

The students were randomly assigned into different groups where one student was given the role of A and the other the role of B. Then the teacher introduced the collaborative learning task to the students. The time allotted to the task was about 20 min. The instruction and materials were delivered in Estonian, and the examples provided here are the authors' translations into English.

This study's data sources included chat messages and recorded actions from the simulation. The simple learning analytics was performed using the Python programming language and the Pandas data analysis library. Regarding qualitative data analysis methods, this study used content analysis methods to analyse data across all dyads. Content analysis is the most frequently used technique when studying collaborative learning processes with chat. Content analysis is a "research method that builds on procedures to make valid inferences from the text" [22].

We developed a coding scheme to analyse and evaluate how students established shared understanding and shared information during their online chats (Table 1). Two

Table 1. Coding scheme used to determine the performance score for groups.

Category	Description	Example
Shared understanding	Student A mentions that their question asks whether it is possible to balance the seesaw with three objects	*"I have to see if it is possible to balance the seesaw using three objects"*
	Student B mentions that their question asks whether it is possible to balance the seesaw with two objects	*"My question is two objects. Is it possible to balance the seesaw using two objects?"*
Information sharing	When the seesaw is balanced, student A mentions the location of their objects on the seesaw and the mass of each one	*"I have 10 kg on the number one."* *"I put a person (30 kg) on 1"* *"I have 5 kg brick on 2 and 10 kg brick on 1"*
	When the seesaw is balanced, student B mentions the location of their objects on the seesaw and the mass of each one	

raters coded all the chats independently based on the coding scheme, and the inter-rater reliability was calculated via the Kappa coefficient. The results demonstrated that the Kappa coefficient was 0.78. The raters discussed and solved discrepancies. The coding scheme calculated the groups' performance scores, where 4 points was the maximum.

4 Results

4.1 Learning Analytics Analysis of Collaborative Performance

The first set of analyses examined the correlation of performance score with the simple learning analytics indicators (Table 2). There were no strong positive correlations between performance score and the length of chat, the number of question marks in the chat, or the time spent on the assignment.

Simply taking count of characters as an analysis unit showed that it is insufficient to differentiate between high and low performing dyads. For example, one dyad (see Table 3) discussed the simulation and assignment but used a lot of slang in their communication. The parts of the text that are irrelevant to solving the task are listed below.

Another example (shown in Table 4) suggests why question marks are not good indicators, since although students may ask substantive questions, they often like to emphasise their question using several question marks together. Alternatively, some students do not use question marks when chatting (see again Table 4).

Chat length in minutes did not correlate strongly with the success of a group. An example shown in Table 6 is of a dyad where the time for solving the task was 20 min. In this example the pair did not start working together before 11 min (Table 5).

Table 2. Correlation matrix of learning analytics indicators.

	Performance score	Chat length	Number of question marks	Time spent
Performance score	1.00	0.35	0.31	0.13
Chat length		1.00	0.73	0.54
Number of question marks			1.00	0.39
Time spent				1.00

Table 3. Example of irrelevant talk that increases the chat length of messages.

09:47	*A: haha*
10:45	*A: hahahahahahah*
15:43	*B: I can't do hahaha anymore*
20:19	*B: AAAAA*
20:23	*A: hahahahaha*
20:24	*B: hahah vb yeah*
20:45	*B: oeki, nice hahha*

Table 4. Examples of distinctive type of communication related to asking questions.

Excessive usage of question marks	Examples of no punctuation
B: so v ??	*A: where do you get these bricks*
B: sooooo ??	*B: what you have there*
A: I guess it was in place ???	*A: which number on where object*

Table 5. Example of student A not participating for a long period of time.

00:00 B	*Heihei!*
10:19 A	*HEii*
10:29 A	*srry, I was doing something else xd*
10:44 B	*All good :D*
11:01 A	*yeez <3*

One dyad who worked together longest was not using the A-side of the seesaw at all; student A thought they succeeded when there was nothing on it.

Table 6. An example of balancing the seesaw without any objects on the seesaw.

A: Balanced	*B: The answer is probably that the seesaw must be empty, then it is balanced*
A: Hurray! What did you do?	*B: I took everything off*

Time pressure was visible from the simulation activity. There were sentences like, "but time is running out," "let's finish; the time is out", and "However, time is running short". This may have prematurely shortened the time on task.

4.2 Qualitative Analysis of Collaborative Performance

A qualitative analysis was carried out to see if there were similarities among all the before mentioned groups and their performance. Three separate groups with similar working patterns were identified based on content analysis of students A and B's chat communication and their actions in modifying the simulation (see Fig. 3).

To solve the problem, students had to understand that they needed to use the chat application to communicate and share information. They needed to understand that seesaw and objects are shared, and to understand that to solve the task successfully, it is necessary to do two parts of the task and share their objects and the locations of objects with a peer. As the chat was the ground for analysing and scoring performance, dyads who communicated outside of the simulation messenger (e.g., using social media messenger) did not achieve high-performance scores even though they may have solved the problem.

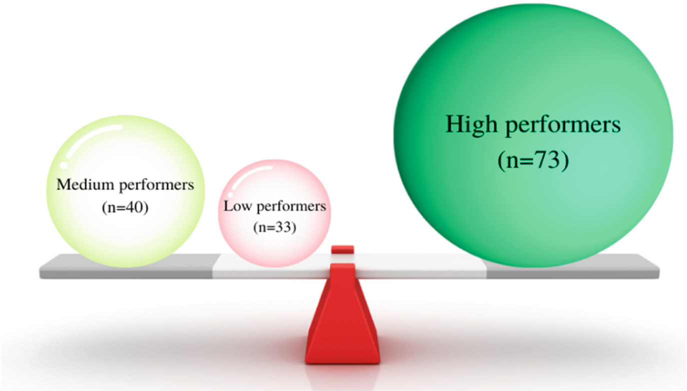


Fig. 3. Three groups were identified based on performance and collaborative behaviours.

Low Performers

Low performers (n = 33) included groups who scored low both on shared understanding and information sharing (0 or 1 point in total performance score). Neither pair of students knew their goal and neither shared information about objects and locations on the seesaw. These groups did not reach a solution.

The following figures illustrate the problem-solving process of different groups (low, medium and high performers). Here, students A and B are distinguished, and the intro phase, the trial-and-error phase, and the answer reporting phase are outlined for both learners. The horizontal axis represents minutes, and the phases are indicated based on the moment they were expressed in the chat discussion (introduction, answer reporting or at which moments the seesaw simulation was manipulated the trial-and-error phase). Examples from the chat application are given that characterise the different phases.

An example of a low performing dyad is shown in Fig. 4 where student B greeted their partner in the first minute, and student A greeted after the second minute. Student B placed the object on their side of the seesaw after 8 min, although for A, it took almost 10 min to put the first element on his side. This dyad was actively experimenting with the simulation but did not find a solution with 2 or 3 objects. A segment of their chat discussion is shown in Table 7 and indicates that they were struggling with balance the seesaw (Table 7).

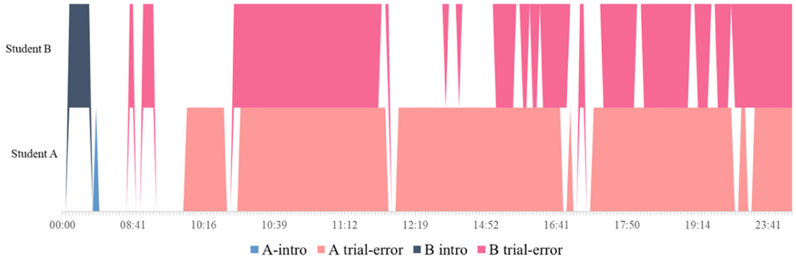


Fig. 4. A low performing dyad characterised by a long time for both students to perform key actions at the start of the activity.

Medium Performers

Medium performers (n = 40) included groups who mainly were successful at information sharing but not at shared understanding. The performance of these groups were at the average level (2 and 3 points) where one or both group members were not aware of the goal, but both or one of the students were/was sharing information. These medium performers did not reach a complete solution. Most of these dyads were solving the student A question - balancing a seesaw with three objects. Of the 40 dyads in this group, only four solved the student B question, primarily due to random trial-and-error to balance the seesaw with two objects. For example, Fig. 5 and Table 8 illustrate a typical group in which they understand that they have to share objects but do not see that they have to find a solution with 2 and 3 objects on the seesaw. They find a balance randomly with three objects and share three locations to end their work.

High Performers

High performers (n = 73) included groups where both members were aware of the goal,

Table 7. Low performer dyad problem-solving discussion.

/…/
06:38 B: do you understand what needs to be done?
07:45 A: not really like not
07:58 B: I guess this seesaw needs to be balanced, but here it looks like it is balanced
08:21 B: something has to be dragged there, some kind of object but I don't really understand it either
08:51 B: i got my object: D
08:54 A: it's not here anymore: P
09:41 B: and now it disappeared, you must have taken it: D
11:02 B: I put mine on #2
11:12 B: try different numbers too
11:51 A: I will try
13:13 B: send me something in the meantime: D
15:23 A: add 10 kg to No. 3
16:49 B: before it was like balance for a while
17:23 B: I removed 10 kg now
19:20 A: 5 and 10 should now go to the other side
19:40 B: I put both but it still doesn't help
22:19 A: this is not normal

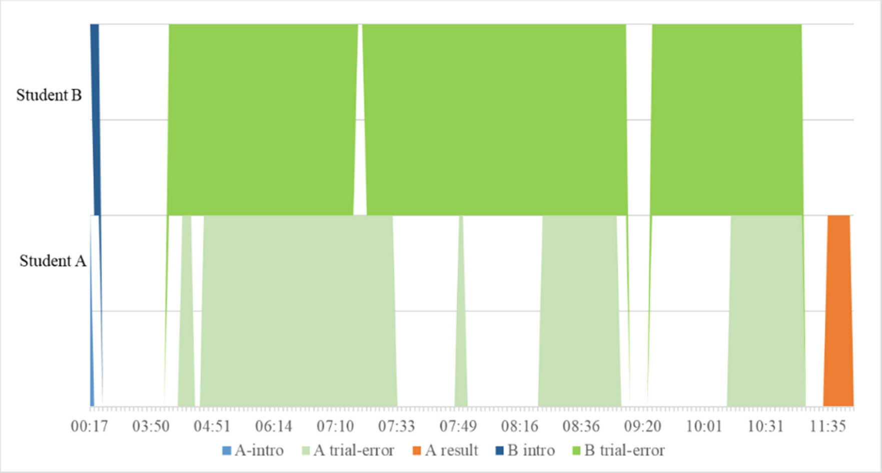


Fig. 5. Medium performer dyad is characterised by usually solving the student A question, even though student B may be strongly engaged in the problem-solving process.

both were sharing information, and most of the groups reached a complete solution. An example of a high performer group is shown in Fig. 6. Both students greeted each other in the first minute, and student A started testing their side of the seesaw. After six minutes, student B also started placing objects on the seesaw in different places. The high performing group is characterised by the fact that a complete solution is found, shown in blue in this figure, answer B at minute 14 and answer A at minute 17.

Table 8. Medium performer dyad problem-solving discussion.

00:17 A: Hi
00:41 B: Hii
02:10 B: how do we work with this model?
02:53 A: I do not know
03:00 B: me neither
03:17 A: oi
03:22 B: I started
04:38 A: we have to answer the question
05:09 A: what is in the bottom
05:16 B: you have to put a man also?
05:24 A: put a man on the seesaw
05:33 B: I did it
05:39 A: I don't see it
05:45 B: I don't see your man too
05:47 B: this has to be like that
05:57 A: ok
06:07 B: how much your man weights
06:08 B: we have to make it balance
06:14 A: I put bricks
06:21 A: 50
06:50 A: I gave you the man
07:00 B: my man weights 30
07:05 A: put something else on the seesaw
07:07 B: I don't have anything
07:09 B: give me the brick
07:10 A: wait
07:22 B: I lost my man
07:33 A: I lost everything
07:34 A: only one brick is left
08:05 A: put 50
08:53 B: put on your side
09:17 A: i lost it again
09:47 B: give me the brick 10
10:11 B: I put 50
10:31 B: put on 1
10:37 B: or on 2
10:51 A: I did it
10:55 B: oo
11:03 B: we did it
11:19 A: we solved
11:30 B: where is your brick
11:31 A: we have to write it
11:35 A: 2
11:37 B: super
11:39 A: where is yours
11:51 B: man is on 3
11:55 B: brick is on 1
11:59 A: good

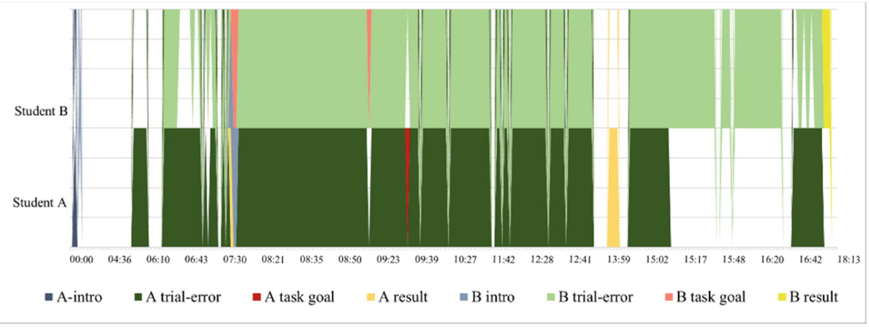


Fig. 6. A high performer dyad is characterised by understanding that they have different task expectations and they balanced the seesaw until they find a full solution.

A prerequisite to finding a complete solution is the understanding that they must balance a seesaw with two and three objects (Table 9). The Fig. 6 illustrates performance with red markers on B - 7th minute and again on 9th minute, and A with red marker on 9th minute.

Table 9. High performance dyad problem-solving discussion.

/.../	*02:47*	*A:*	*Do we have the same weight objects? I have a 30 kg person and three red blocks of different weights.*
	03:04	*A:*	*50 kg, 10 kg and 5 kg*
/.../	*04:20*	*A:*	*I have these "objects" under this seesaw and then on the right a box "drag the object there to share it"... maybe I have to share them with you?*
	04:25	*B:*	*to drag the object here to share it*
/.../	*06:10*	*B:*	*I guess the easiest to do is to try so that 50 kg is somewhere around one*
	06:15	*A:*	*try to put both on 4 if you can*
/.../	*07:22*	*B:*	*did we only need to use two?*
	07:26	*B:*	*as it is in the description*
	07:28	*A:*	*three*
	07:52	*B:*	*it is written we have to use together 2 objects*
	08:00	*B:*	*or you have another*
	09:13	*B:*	*I have in the description that two objects need to be used*
	10:08	*B:*	*shall we try with three?*
/.../	*13:17*	*B:*	*aa maybe we have to do this to try with two objects*
	13:20	*B:*	*and then with three objects*
	13:28	*B:*	*what objects are you using now*
	13:55	*A:*	*then write down in which case it worked with two*
	13:55	*A:*	*I have a person alone now on 1 again*
	13:55	*A:*	*I guess we have to do it separately for both questions*
	13:55	*A:*	*we got the answer to your question*
	13:55	*A:*	*I had a person [30kg] on 1*
	13:55	*A:*	*what are you on seesaw now?*
	13:55	*A:*	*I have 10 and a person*
	13:55	*A:*	*with three is ok*
	14:24	*B:*	*I had man on 1 and I don't know what you had; let's try with two*
/.../	*16:56*	*B:*	*where do you have a person*
	17:04	*A:*	*on 1*
	17:37	*B:*	*ok done then?*

5 Discussion

Our first research question asked how simple learning analytics indicators relate to collaborative performance on an asymmetric digital task. In short, the analysed indicators had weak positive correlations with performance score. The number of chat characters was not a good indicator because it did not differentiate between high and low-performance dyads. Some groups talked a lot in the chat, but it was not relevant to solving the task; other groups used slang and repeated the same characters in the chat. Previous studies [17] show that low-collaborating groups are observed to act in parallel without discussing, whereas high-collaborating groups collaborate on task-related objects while discussing. However, in our study, it was not possible to differentiate between groups based on chat length.

The number of question marks was also not a good indicator because some dyads did not use question marks in their written text when asking questions. Some dyads used several question marks in the same question. As there was a time limit for the assignment, the time spent on the assignment was not a good indicator for group score. There was no correlation between the longest or shortest time on the task with the group score. Furthermore, time pressure was visible from the simulation activity.

The second research question aimed to find the relationship between qualitative content analysis of the collaborative process and performance on the asymmetric digital task. By coding the messages and evaluating the actions to manipulate the simulation, we were able to identify three unique groups: low, medium, and high performers. Analytics of task-specific activities when learners collaborate in complicated problem-solving contexts have previously been demonstrated in research to identify high and low performers in collaborative learning [23–25].

As the problem-solving task was novel for students, we found similar results as in a study by Järvela et al. 2016, where students may have had difficulty seeing the task from a different perspective [26]. As a result, in our study, student B was in an adverse position because of their initial lack of objects to place onto the seesaw and the dependence on student A to share at least one object. When one student has unique information or resources, but does not share it, dyads tend to fail.

Some dyads in our study could not balance the seesaw because they were trying to place objects on it simultaneously. A prior study found that when participants touch unrelated objects on the screen and several users engage with the screen simultaneously, it may negatively predict cooperation quality [16].

During the beginning of this task, there was apparent uncertainty among the students. In the context of collaborative learning, the dynamic and reciprocal adaptation of shared interaction occurs; that is, when individuals in a group are not only engaged in the same task at the same time but are also cognitively "in tune" [19, 27]. Dyads modifying the simulation and sharing information overcame their early phase. Learners who collaborate to solve real, open-ended tasks tend to fail, which is beneficial for learning since it leads to deep cognition and increased transfer [24]. There is a "zone of optimum confusion" [28] where learners become aware of their knowledge gaps and then realise the profound aspects of the underlying notion when they work on challenging tasks [29]. Confusion might be beneficial in this zone. If learners' uncertainty persists, it might become ineffective, leading to frustration and disengagement [30]. According to one study, effective teams better reflect the issuing state by concentrating on the interdependence of the factors critical to accomplishing the task goals. Less effective teams, on the other hand, tend to focus on connections between factors that have a minor impact on meeting task objectives [31].

We found that indicators of the high-performance group were that both students used the chat for communication and sharing information. They understood that the seesaw and object were shared between them. Students had to solve both parts of the task, and they needed to share where and which masses were on the seesaw.

The asymmetric simulation used in the study is an excellent opportunity to develop collaboration skills together with subject skills. Although this was presumably the first experience for students with this type of problem-solving task, half of the dyads were successful. These results suggest that even a 20-min collaborative task can provide a meaningful collaborative learning experience. A list of several more collaborative simulations can be found at https://leosiiman.neocities.org/simulations.html.

The chat application mode of communication used in this study made learning visible. For some learners, this was inconvenient and unfamiliar, and they would have preferred face-to-face communication. If a chat application is used to make learning visible, the importance of using it should be better explained to learners.

We must consider that group performance does not directly reflect the learning gain. According to one recent study [21], it was found that among learners who achieved more significant learning gains, some teams failed the assignment, while others succeeded but did not learn.

Finally, it is important to remark that this study was a one-time intervention. In general, several lessons with a similar structure should be designed to help learners better understand the logic of an asymmetric collaborative scenario. Our findings indicate that information sharing and developing shared understanding needs to be practised, which is facilitated by students learning the structure of an asymmetric collaborative task and focusing more on collaborating to solve the problem instead of on technical and organisational problems.

6 Conclusion

Future studies on asymmetric collaboration should consider new approaches to learning analytics, such as voice recognition or natural language processing (NLP), that can better identify relevant collaborative behaviours in computer-supported collaborative learning [32]. Furthermore, automated content analysis is one solution to support a lecturer in providing guidance and feedback to students [33]. The automated algorithm must handle natural language input in this task and determine whether pairs have established shared goals.

We should remember that some researchers have found neither alignment of learner actions nor learners building on each other's reasoning was connected to task performance in a chat-based collaborative learning environment. Other elements, such as group dynamics and prior knowledge, had a more significant influence [19]. We recommend that students practice with asymmetric collaborative activities in order to develop their collaboration problem-solving skills.

A recent study [34] found that when co-regulation was familiar in a group, students with less monitoring accuracy obtained better learning, as predicted by theory. The findings revealed that complex interaction of individual and group-level variables influencing collaborative learning outcomes. This should be addressed while developing measures to support monitoring. For example, prompts, identified as a critical strategy to promote metacognitive engagement [32], may be needed as support early when students collaborate to solve an asymmetric digital task.

Our findings align with Tang et al., 2022 results [35]: First, students must pay close attention to online collaborative activities since attentiveness is a favourable indication of knowledge gains in educational contexts. Second, teachers or intelligent systems must monitor student collaborative behaviours and intervene when groups tend to work alone. This style of collaboration is unproductive because learners are not focused on learning or collaboration. Third, students in the distributed collaboration model may become distracted as time passes. Furthermore, if their attention is drawn away from the collaborative activity, learners in this mode find it challenging to return to the full attention mode [35].

A limitation of our study is that reliability and validity are issues when the content analysis method is applied. Therefore, the results of our coding schema to other contexts may not generalise.

References

1. Zeichner, K.: The turn once again toward practice-based teacher education. J. Teach. Educ. **63**(5), 376–382 (2012)
2. Fiore, S.M., Graesser, A., Greiff, S.: Collaborative problem-solving education for the twenty-first-century workforce. Nat. Hum. Behav. **2**(6), 367–369 (2018)
3. OECD: PISA 2015 Results (Volume V): Collaborative Problem Solving. PISA, OECD Publishing, Paris (2017)
4. Griffin, P., Care, E. (eds.): Assessment and Teaching of 21st Century Skills. Springer, Dordrecht (2015). https://doi.org/10.1007/978-94-017-9395-7
5. Johnson, D.W., Johnson, R.T.: Making cooperative learning work. Theory Pract. **38**(2), 67–73 (1999)
6. Siiman, L.A., Rannastu-Avalos, M., Mäeots, M.: Developing smart device friendly asymmetric simulations for teaching collaborative scientific inquiry. In: 20th International Conference on Advanced Learning Technologies (ICALT), pp. 130–131. IEEE (2020)
7. Siiman, L.A., Rannastu-Avalos, M., Mäeots, M., Pedaste, M.: The Go-Lab ecosystem: a practical solution for school teachers to create, organize and share digital lessons. Bull. Tech. Committee Learn. Technol. **20**(2), 27–35 (2020)
8. Rannastu, M., Siiman, L.A., Mäeots, M., Pedaste, M., Leijen, Ä.: Does group size affect students' inquiry and collaboration in using computer-based asymmetric collaborative simulations? In: Herzog, M.A., Kubincová, Z., Han, P., Temperini, M. (eds.) ICWL 2019. LNCS, vol. 11841, pp. 143–154. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-35758-0_14
9. Roschelle, J., Teasley, D.: The construction of shared knowledge in collaborative problem solving computer-supported collaborative. Learning **128**, 69–97 (1995)
10. Järvelä, S., Järvenoja, H., Malmberg, J.: Capturing the dynamic and cyclical nature of regulation: methodological Progress in understanding socially shared regulation in learning. Int. J. Comput.-Support. Collab. Learn. **14**(4), 425–441 (2019). https://doi.org/10.1007/s11412-019-09313-2
11. Barron, B.: When smart groups fail. J. Learn. Sci. **12**(3), 307–359 (2003)
12. Martinez, R., Wallace, J.R., Kay, J., Yacef, K.: Modelling and identifying collaborative situations in a collocated multi-display groupware setting. In: International Conference on Artificial Intelligence in Education, pp. 196–204 (2011)
13. Reilly, J.M., Schneider, B.: Predicting the quality of collaborative problem solving through linguistic analysis of discourse. In: Proceedings of the 12th International Conference on Educational Data Mining (EDM), EDM 2019, pp. 149–157 (2019)
14. Viswanathan, S.A., VanLehn, K.: Using the tablet gestures and speech of pairs of students to classify their collaboration. IEEE Trans. Learn. Technol. **11**(2), 230–242 (2017)
15. Weinberger, A., Fischer, F.: A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. Comput. Educ. **46**(1), 71–95 (2006)
16. Evans, A.C., Wobbrock, J.O., Davis, K.: Modeling collaboration patterns on an interactive tabletop in a classroom setting. In: Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW, vol. 27, pp. 860–871 (2016)

17. Martinez-Maldonado, R., Dimitriadis, Y., Martinez-Monés, A., Kay, J., Yacef, K.: Capturing and analyzing verbal and physical collaborative learning interactions at an enriched interactive tabletop. Int. J. Comput.-Support. Collab. Learn. **8**(4), 455–485 (2013). https://doi.org/10.1007/s11412-013-9184-1
18. Nasir, J., Kothiyal, A., Bruno, B., Dillenbourg, P.: Many are the ways to learn identifying multi-modal behavioral profiles of collaborative learning in constructivist activities. Int. J. Comput.-Support. Collab. Learn. **16**, 485–523 (2022). https://doi.org/10.1007/s11412-021-09358-2
19. Popov, V., van Leeuwen, A., Buis, S.: Are you with me or not? Temporal synchronicity and transactivity during CSCL. J. Comput. Assist. Learn. **33**(5), 424–442 (2017)
20. Rodríguez, F.J., Boyer, K.E.: Discovering individual and collaborative problem-solving modes with hidden Markov models. In: Artificial Intelligence in Education: Proceedings of the World Conference on AI in Education 2015, pp. 408–418 (2015)
21. Hadwin, A.F., Järvelä, S., Miller, M.: Self-regulated, co-regulated, and socially shared regulation of learning. In Zimmerman, B.J., Schunk, D.H. (eds.) Handbook of Self-regulation of Learning and Performance, pp. 65–84. New York, vol. 12, pp. 8–22 (2001)
22. Rourke, L., Anderson, T., Garrison, D.R., Archer, W.: Methodological issues in the content analysis of computer conference transcripts. Int. J. Artif. Intell. Educ. (IJAIED) **12**, 8–22 (2001)
23. Emara, M., Rajendran, R., Biswas, G., Okasha, M., Elbanna, A.A.: Do students' learning behaviors differ when they collaborate in open-ended learning environments? In: Proceedings of the ACM on Human-Computer Interaction, vol. 2, no. CSCW, pp. 1–19 (2018)
24. Kapur, M.: Temporality matters: advancing a method for analyzing problem-solving processes in a computer-supported collaborative environment. Int. J. Comput.-Support. Collab. Learn. **6**(1), 39–56 (2011)
25. Perera, D., Kay, J., Koprinska, I., Yacef, K., Zaïane, O.R.: Clustering and sequential pattern mining of online collaborative learning data. IEEE Trans. Knowl. Data Eng. **21**(6), 759–772 (2008)
26. Järvelä, S., et al.: Socially shared regulation of learning in CSCL: understanding and prompting individual- and group-level shared regulatory activities. Int. J. Comput.-Support. Collab. Learn. **11**, 263–280 (2016)
27. Baker, R.S., D'Mello, S.K., Rodrigo, M.M.T., Graesser, A.C.: Better to be frustrated than bored: the incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. Int. J. Hum. Comput. Stud. **68**(4), 223–241 (2010)
28. Lodge, J.M., Kennedy, G., Lockyer, L., Arguel, A., Pachman, M.: Understanding difficulties and resulting confusion in learning: an integrative review. Front. Educ. **3**, 1–10 (2018)
29. Loibl, K., Rummel, N.: The impact of guidance during problem-solving prior to instruction on students' inventions and learning outcomes. Instr. Sci. **42**(3), 305–326 (2013). https://doi.org/10.1007/s11251-013-9282-5
30. D'Mello, S., Graesser, A.: Dynamics of affective states during complex learning. Learn. Instr. **22**(2), 145–157 (2012)
31. Dindar, M., Järvelä, S., Nguyen, A., Haataja, E., Çini, A.: Detecting shared physiological arousal events in collaborative problem solving, Contemp. Educ. Psychol. **69**, 102050 (2022). ISSN 0361-476X
32. Malmberg, J., Järvelä, S., Järvenoja, H.: Capturing temporal and sequential patterns of self-, co-, and socially shared regulation in the context of collaborative learning. Contemp. Educ. Psychol. **49**, 160–174 (2017)
33. Alvarez, C., Zurita, G., Carvallo, A., Ramírez, P., Bravo, E., Baloian, N.: Automatic content analysis of student moral discourse in a collaborative learning activity. In: Hernández-Leo, D.,

Hishiyama, R., Zurita, G., Weyers, B., Nolte, A., Ogata, H. (eds.) CollabTech 2021. LNCS, vol. 12856, pp. 3–19. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-85071-5_1

34. Haataja, E., Malmberg, J., Dindar, M., Järvelä, S.: The pivotal role of monitoring for collaborative problem solving seen in interaction, performance, and interpersonal physiology. Metacogn. Learn. **17**, 241–268 (2022)

35. Tang, H., Dai, M., Yang, S., Du, X., Hung, J.L., Li, H.: Using multimodal analytics to systemically investigate online collaborative problem-solving. Dist. Educ. **43**, 1–28 (2022)

Using Process Mining Techniques to Discover the Collective Behaviour of Educators in a Learning Community Platform

Oscar Esteban Bajo(✉), Ishari Amarasinghe, Nicolás Felipe Gutiérrez-Páez, and Davinia Hernández-Leo

ICT Department, Universitat Pompeu Fabra, Barcelona, Spain
oscar.esteban01@alumni.upf.edu, {ishari.amarasinghe, nicolas.gutierrez,davinia.hernandez-leo}@upf.edu

Abstract. Learning design platforms aim to facilitate teachers with their learning design processes. Existing studies have mostly focused on developing design features for learning design platforms yet the number of studies that investigate the collective behaviour of educators when using such tools is scarce. To this end, this study proposes the use of data analytics techniques, namely, process mining, to analyse the behaviour of teachers as they engage in using an online learning design platform as part of a teacher professional development course. The findings of the study shed light on teachers' collective behaviour and motivational aspects related to their participation within an online learning design community.

Keywords: Learning design · Learning communities · Community of educators · Process discovery

1 Introduction

Learning Design (LD) can be understood as the field of study that aims to explore how educators prepare and revise a set of learning activities toward more pedagogically informed decisions to achieve particular educational objectives [1,2]. Research in learning design recognises *teachers as designers* of their own learning situations and aims to facilitate educators in collectively building and sharing their design decisions and experiences [1,3]. Over time, different online platforms to support learning design processes have been developed. Some examples for such platforms include Learning Designer [4], Cloudworks [5], Learning Activity Management System (LAMS) [6], Graasp [7] and Integrated Learning Design Environment (ILDE) [8]. The aforementioned online platforms support the learning design process at different stages from conceptualising, authoring and implementing, and even providing the possibility of sharing and co-designing learning designs within teacher communities [8]. These types of learning design environments are commonly used to support formal teacher training initiatives (e.g.

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 175–189, 2022.
https://doi.org/10.1007/978-3-031-20218-6_12

being used in professional development courses) or teachers' informal learning interests (e.g. searching for inspiration), and thus become also learning community platforms [9].

Different data-driven approaches can be used to analyse teachers' collective behaviour when using learning design community platforms. Such an analysis could provide insights towards educators' collective knowledge building, e.g., requesting and sharing new knowledge from colleagues, therefore how social learning occurs in online learning design platforms [7]. Moreover, knowledge gained from *community analytics for learning design* [10], can provide insights such as which designs to reuse based on their popularity, or with whom to collaborate on design tasks and which features provided by those platforms require improvements.

Despite the wealth of insights such analysis could bring, the number of studies that analyse the behaviour of teachers in online learning design communities as well as their motivational factors that enhance active participation in such communities is limited [10,11]. This study aims to address this gap in existing research as a proof of concept, and in particular, we address the following research question: *What is the behaviour of educators when using an online learning design platform during teacher professional development Massive Open Online Courses (MOOCs)?* Answering the aforementioned research question would contribute not only to understand the educators' behaviour in a teacher community platform for learning design but also will inform on participants' drivers and motivations as well as may suggest ways to enhance the platform's design.

2 Background

The possibility to collect big amounts of educational data in the digital age allows for large-scale analysis and therefore provides opportunities to obtain insightful information on a multitude of teaching and learning processes. For instance, MOOCs facilitate a large audience of students to learn online, and the digital traces they left behind can be collected and analysed to better understand how they engaged in learning processes and platform features that facilitate better participation and learning [12].

Learning Analytics (LA) constitutes an emerging research area and is defined as the "measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" [13]. LA constitute a wide range of disciplines and analytical techniques, such as machine learning, data mining, information visualisation, and psychology. With the aid of such LA techniques, participants' interactions in online learning environments can be analysed and studied, producing important knowledge on a variety of topics, including participation trends and community involvement [14].

Specifically, the direction we are interested in is process mining, a LA technique that allows to extract knowledge of real-life processes from a registered set of events [15]. On one hand, previous studies have mainly used process mining to investigate aspects related to learning processes, for instance, students learning strategies [16], behavioural [17] and interaction patterns [18]. On the other hand, previous studies on teachers' motivations to engage in collaborative learning design platforms focus on using subjective questionnaires e.g., self-reported motivational factors collected through questionnaires [19] or interviews [20,21]. In contrast, the goal of our study is to use LA techniques, in particular, process mining in order to obtain insights regarding the behaviour of teachers in an online learning design platform.

3 Methodology

To answer the research question, we have followed the case study research method [22]. Case studies are appropriate when the study aims to explore contemporary phenomena or issues in natural settings [23]. The selected case study is a MOOC for teacher training and the learning activities completed using the learning design platform ILDE. The data log was extracted for two runs of the MOOC. The following subsections provide details about the functionalities of the learning design platform, participants of the MOOCs, data collected, and the pre-processing needed. It is important to notice that the term *learning design* will be used to express a single design created by an educator in the platform.

3.1 Platform Functionalities

ILDE is a platform that supports teachers to engage in the full learning design cycle, from conceptualisation to authoring and deployment (see Fig. 1) [8]. By enabling teachers to share their learning designs with the community and to view and comment on each other's designs, the platform also fosters social features [24].

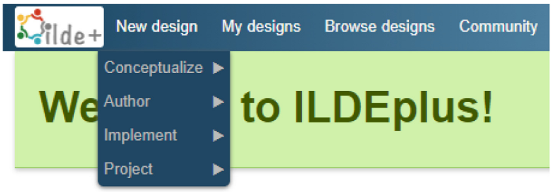


Fig. 1. User Interface of the ILDE platform showing options available under *New design* menu.

Functionalities such as creating, publishing, viewing, and commenting are available, as well as sharing, editing or deleting. As shown in Fig. 1 upper menu allows to access the learning design creator, where each part of the learning design cycle is available. The platform allows to publish learning designs that are completed and enables to share designs publicly with the other members of the community (*Browse designs* tab). A learning design can be viewed, commented by any user (including the creator) and edited or deleted by the creator. An interesting feature is the possibility to create a learning design by duplicating an existing design (by the original creator of the learning design or other users). This is a way to take inspiration from other educators and it gives credit to the original creator, marked on the learning design itself. A learning design's level of popularity can be gauged by counting the references to duplicates [25].

3.2 MOOC Participants

The MOOC was made available to teachers who are interested in online learning. The demographic data such as age, gender, and years of teaching experiences were not collected at the time of the MOOC, as it was not needed. Additionally, there was no participant overlap between the two MOOC runs, indicating that each MOOC run had a distinct group of educators. The total number of participants were 325 and 399, respectively.

3.3 Data Collection

ILDE was used during the MOOCs to facilitate teachers get familiar with the learning design cycle. The MOOC course lasted for 5 weeks (in both runs). Details about weekly activities proposed in both MOOCs are presented in Table 1. The learning design platform was used in all weekly activities of the MOOC except for the activities proposed in the fourth week. In order to conduct a thorough analysis of educators' behavior when using an online learning design platform during a teacher professional development MOOC, we used log data obtained from both MOOC runs separately in this study. After analysing the structure of the data set, we chose a set of actions (that are related to learning design tasks) that would be considered for the subsequent data analysis (see Table 2).

It is important to remark that educators were learning about Learning Design and ILDE alongside the course of the MOOC, and so, findings are not necessarily generalizable for other cases in which educators already knew about the features and use of ILDE.

3.4 Data Pre-processing

We created an event log to apply process mining techniques. An event log is an ordered set of instances that include a user ID, an action, and a point in time [26].

After the event log was prepared, a data cleaning process was carried out. First, all actions performed before the starting date of the MOOC were removed, as only moderators used the platform during this period for testing purposes. Also, events with the exact same timestamp were rearranged. This was a matter

Table 1. Activities proposed for each week of the MOOC

Week	Activities
1	- Design Studio Journal - Dream Bazaar - Convergence session
2	- Get familiar with persona concept - Create your own persona - Analysing context - Objective of your learning activities - Revisit your dream - Convergence session
3	- Search for other learning activities - Define the heuristics - Learn about scenarios - Create scenario - Convergence session
4	- Prototype your artefact - Test your prototype - Consolidate your prototype - Convergence session
5	- Publish your learning activity - Peer feedback - Convergence session

of the structure of the tables in the database, as different activities were gathered from different tables, and some activities naturally came after others (e.g., viewing after logging in). The size of the event log and the user count can be seen in Tables 3 and 4.

Then we choose two algorithms based on their distinct advantages [27] and considering the nature of our data: Heuristics miner and pMineR. Heuristics miner [28] provides an easy-to-read graph that contains starting and end points representing the first and last actions performed by a user. One of the most characteristic features is the use of the dependency metric, which categorises each link connecting two actions from the dependency relationship. However, a limitation of this algorithm is that it only supports single loops, so if an action loops more than once, this information will not be shown in the diagram. pMineR [29] is a process discovery algorithm meant to create a more reading-friendly diagram. It is based on first-order Markov model and shows the transition probability between actions. To generate the process maps we used the pm4py[1] Python package and the bupaR[2] R package.

[1] www.pm4py.fit.fraunhofer.de/.

[2] www.bupar.net/.

Table 2. Used actions.

Action	Description
create	Creating a learning design from scratch
create from duplicate	Creating a learning design by duplicating an existing learning design (this could be a self-duplicate meaning the design was previously created by the same user or a duplicate of someone else's design)
delete	Deleting a learning design
viewed_lds	Viewing a learning design. The user must click on a learning design for this action to be quantified
viewed_profile	Viewing other users' profiles. The user must click on other user's name to view profile
revised_docs	This action captures editing a learning design
generic_comment	Adding a comment to an existing learning design (comment on your own designs or someone else's design)
login	Introducing credentials to access an account
publish	After creating a learning design, it must be published. Published designs are visible to the entire community
share_add_editor	Giving other users permission to edit a learning design
share_add_viewer	Giving other users permission to view a learning design
share_del_viewer	Remove permission to view a learning design
share_add_acv	Making a learning design available to anyone to see in the community using "All Can View" option
share_remove_acv	Removing "All Can View" option from a learning design

Table 3. Size of the event log before and after cleaning.

	run 1	run 2
Original event log	48823	47623
Clean event log	43984	45369

Table 4. Number of active users before and after cleaning.

	run 1	run 2
Original event log	325	399
Clean event log	300	383

4 Results

In the following, we present the results obtained using Heuristics miner and pMineR.

4.1 Heuristics Miner

The models obtained from Heuristics miner considering teachers' actions in the learning design community platform across the two MOOC runs are shown in Fig. 2 and Fig. 3, and related details are given in Table 5 and Table 6 in

Appendix A. The diagrams includes the total event count in brackets and nodes were coloured in different shades based on the event count (e.g., a larger amount of events resulted in a node with a darker shade). Arrows show the direction of the actions with a number indicating the transition count.

Results obtained (Fig. 2 and Fig. 3) show that *login* is the first common action in the design platform alongside with *view*, even after the event log cleaning process (see Table 3 and Table 4). The reason why this should not happen is that these diagrams model the whole flow of actions for each user from the very first action after the green starting point to the last one recorded before the ending point. This might be due to the Heuristics miner's dependency relationship mentioned earlier.

The results further indicate that some actions are sequential, i.e., only followed after performing another action. For instance, *comment*, *create from a duplicate* and *editing* are always preceded by *view*, as the users need to view the learning designs first in order to perform the subsequent actions.

From almost 30000 recorded *view* actions in run 1, in more than 21000 occasions, or 70% of the time, the action was performed consecutively, i.e., the same action was repeated (Fig. 2). In the second MOOC run we obtained similar results. This does not only mean that the *view* action is the most performed one, but if a teacher views a learning design, he or she is very likely repeat the same action. Not only the *view* action was repeated over and over again but similar behaviour was observed regarding the *create* and *delete* actions as well. For instance, the *create* action in the second MOOC run was performed again 1851 times from a total of 2612 (70% of the total count), and the *delete* action was performed 335 from a total of 554, which is a 60% (Fig. 3). This indicates that not only *create* was a more popular action in the second MOOC run, but many learning designs were created in a row, probably because of the incentive of the guidance for this particular MOOC.

4.2 pMineR

pMineR provides diagrams (see Fig. 4 and Fig. 5 and related details in Table 7 and 8 in Appendix A) with different links showing likelihood. Links with a probability equal or under 0.01 are categorized as noise and are not shown. Unlike before, only a few connections are strongly highlighted. Many of these highlights are from connections with a lot of presence previously seen in the Heuristics miner. For instance, the *login* → *view learning design* is shown to be the most likely transition in both MOOC runs. Transitions towards the ending point have a small probability, all of them smaller than 0.01 in the case of the first run.

The first observation from the figures is that there is only one action after the starting point towards the *login* action. This confirms what we stated in the previous subsection. The pMineR model does not use the same process to generate the diagrams as the Heuristics miner, but a likelihood estimation based on the sequence count. This means that a probability of 1 towards the *login* action states that the totality of users had this action as their first.

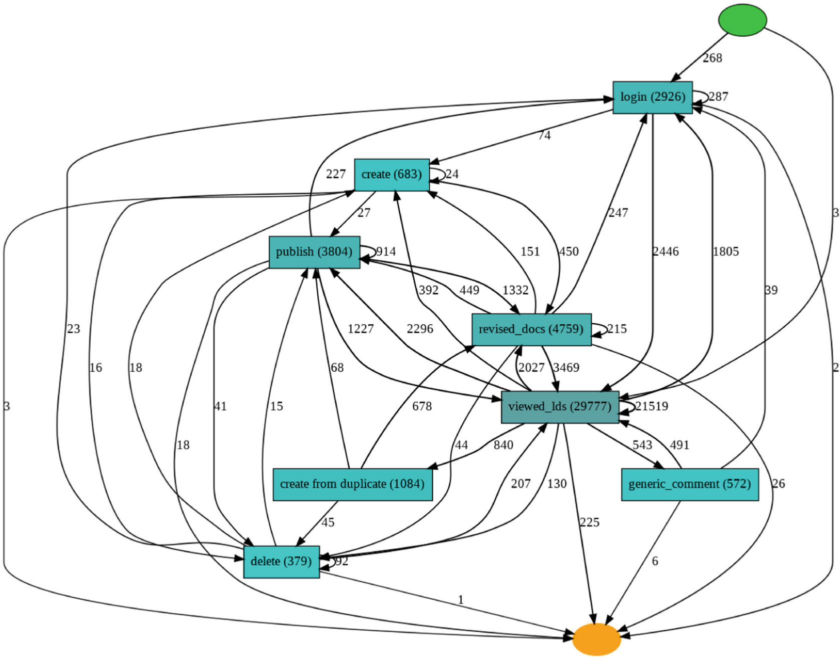


Fig. 2. Heuristics net in MOOC run 1.

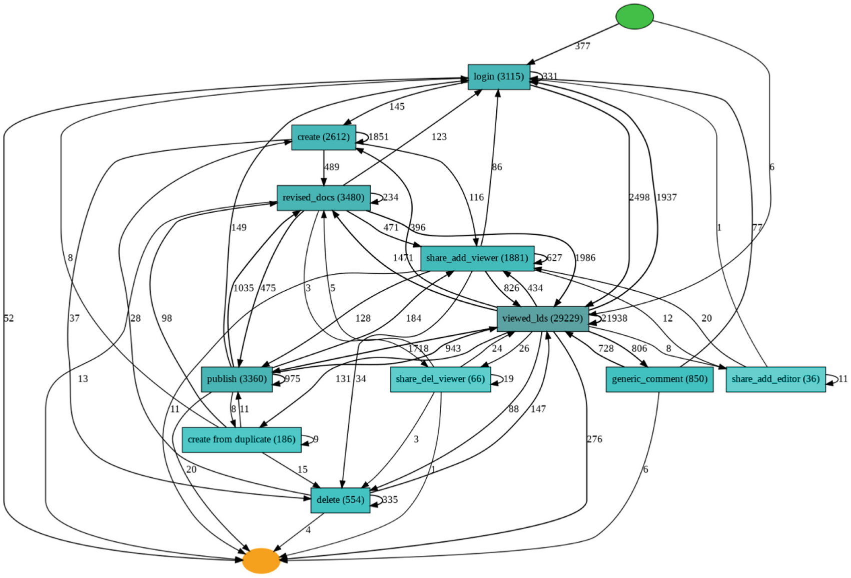


Fig. 3. Heuristics net in MOOC run 2.

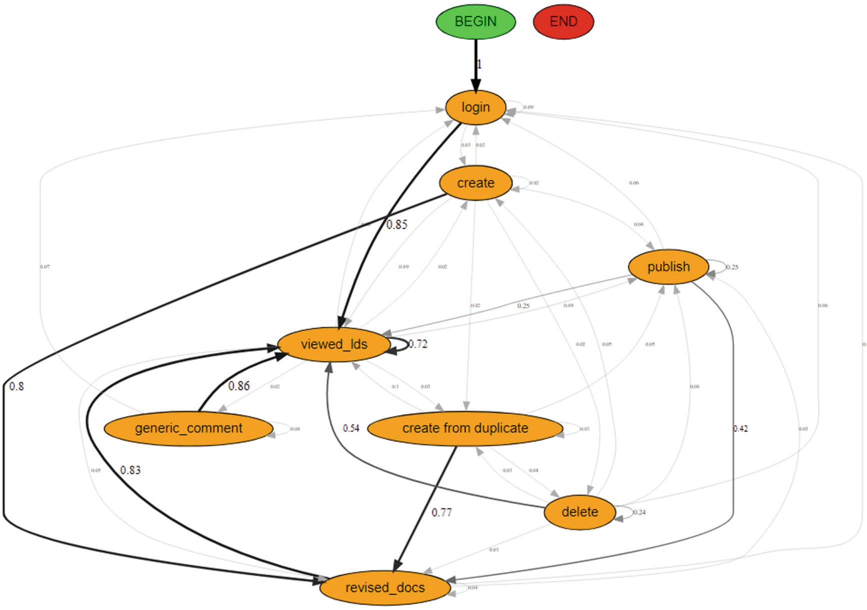


Fig. 4. pMinerR with a 0.01 threshold in MOOC run 1.

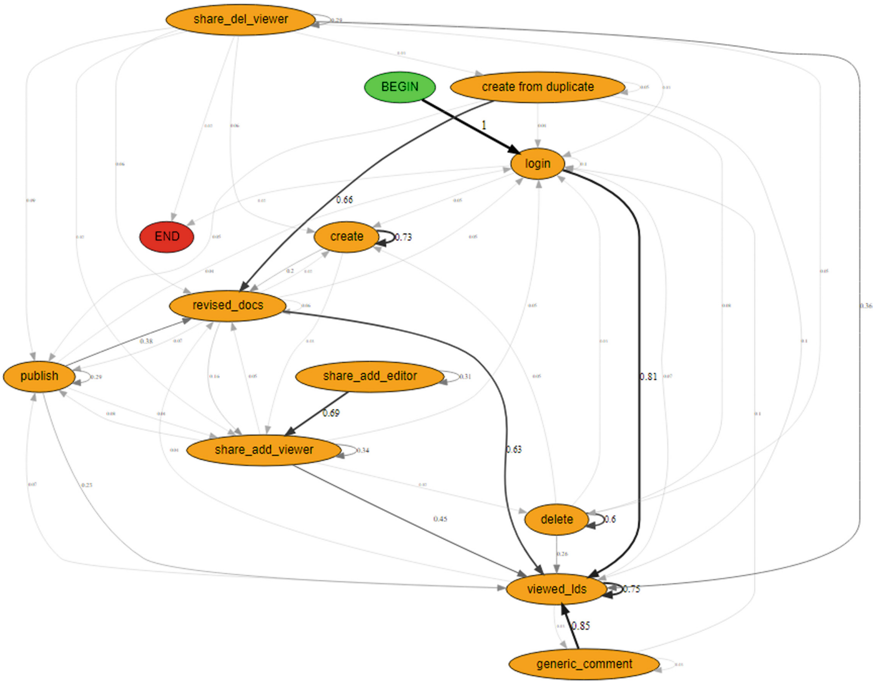


Fig. 5. pMinerR with a 0.01 threshold in MOOC run 2.

With pMineR it is also visible that some actions are always followed by other actions. Nevertheless, this diagram shows weaker connections for the aforementioned examples. Additionally, it can be seen that in the first MOOC run, *comment* → *view* is the path with the highest probability. All probabilities of paths from a given node sum up to one, so here, probabilities are mainly useful to assess the most popular action from a given node.

Finally, the likelihood of succeeding actions is well displayed with pMineR, and the same percentages mentioned in the Heuristics miner can be extracted as well in the pMineR diagrams.

5 Discussion

Within the research question defined for this study we aimed to understand what is the behaviour of educators when using an online learning design platform during teacher professional development MOOCs. It is important to note that the main objective of this study was not to arrive at generalisable results, but rather to give the first insights into the behavioral patterns of educators within a learning design community platform. We looked at the general behaviour of educators when using an online learning design platform during teachers' professional development MOOCs.

According to the study's findings, educators spent the majority of their time investigating other's learning designs in the particular learning design environment. This may indicate that educators are curious to explore other's learning designs when they are in line with their own pedagogical intentions and goals. These results also show that the community exploration was made feasible by the features of the learning design platform, and educators were able to become familiar with the platform and use its features throughout the professional development course.

The fact that the number of *edit* actions is always higher than the number of *create* actions (see Fig. 2 and Fig. 3) suggests that educators were interested not only in creating designs but also in improving them. This result can be interpreted as an exploration of others' learning designs that may have resulted in a rich knowledge-building experience [7] that enabled educators to learn from other's designs and to incorporate the new knowledge into their designs.

It was also prominent the presence of repeated actions, especially *viewing learning designs* and *create* in the second MOOC run. These findings also imply that MOOC design encourages participants to explore and develop learning designs in order to enhance their skills and knowledge, which is consistent with previous research findings on educators' motivations to participate in online communities [19–21]. The findings also showed that since the actions that promote social interaction among participants (comment, share and co-create) are less frequent and are only promoted during the last week of the MOOC, it is necessary to refine and improve the MOOC design as well as functionalities of the learning design platform to further enhance the social learning experience within the community.

Regarding the chosen algorithms, the Heuristics miner provided better and more explanatory results for our data set when compared to the other approaches. The graphical results obtained using pMineR are similar to Heuristics miner in the sense that they also displayed start and ending points and arrows between actions showing direction. Additionally, the thickness of the arrows was modified to better represent the transition probabilities. Almost all conclusions that were taken from the Heuristics miner could be taken from pMineR as well. A difference is that the starting point only followed the login action, which made more sense after the data cleaning process.

As previously stated, studies on educators' behaviour in learning design platforms is a very underdeveloped area. Hence, as a result, it became difficult to compare our study to previous similar investigations.

There are several limitations of this study. First, we encountered limitations during the selection of actions to consider. For instance, a logout action was seen as relevant to be considered when creating the event logs, however, due to technical issues some logout actions were not recorded in the data set. Hence, we decided to discard such actions when preparing the data set for the subsequent analysis. If it had been available, process mining diagrams could have represented all sessions from a login to a logout for all users instead of the whole flow of actions from beginning to end. Second, the study itself reflects the use of a learning design environment by the educators while being taught how to use it during a MOOC course which greatly affects their use of the platform and consequently limits the insights that can be drawn about their regular use of the platform.

6 Conclusions and Future Work

In this study, we have used two process mining techniques to analyse educators' behaviour in a platform that supports the full cycle of learning design. The educators were given a set of activities to perform on the platform, which was conducted in the context of a MOOC to learn about the concept of Learning Designs.

The findings of the study revealed interesting observations about educators' behaviour within this learning design community platform. Process discovery trees provided the most informative and interpretable representations regarding educators' behaviour, e.g., the Heuristics miner, with features such as the total event count, as well as the transition direction and transition count.

Future work includes the categorization of educators into separate groups before the analysis of their actions. This could be done using machine learning techniques such as K-means [30]. After the categorisation, independent analyses could be done on each group, which would result in more focused results in comparison to a single analysis for the whole users' log. Moreover, the current study findings might be expanded to investigate whether the educators' behavior will change in response to the various types of activities suggested for each week of the MOOC course.

Additionally, comparing how educators behave collectively in various contexts, such as how they use the same ILDE platform in a MOOC that is not necessarily framed in teacher professional development courses, could reveal additional insights into the existence of common patterns of use that would eventually facilitate to produce thorough and generalisable insights into how educators behave in learning design platforms and across contexts.

Acknowledgements. This work has been partially supported by the grants PID2020-112584RB-C33 funded by MICIN/AEI/10.13039/501100011033. D. Hernández-Leo (Serra Húnter) acknowledges the support by ICREA under the ICREA Academia programme.

A Matrix Form for Process Maps

Table 5. Matrix form of Fig. 2

	login	create	create from d.	publish	revised docs	viewed lds	comment	delete	END
START	268	0	0	0	0	32	0	0	0
login	287	74	0	0	0	2446	0	0	21
create	0	24	0	27	450	0	0	16	3
create from dup.	0	0	0	68	678	0	0	45	0
publish	227	0	0	914	1332	1227	0	41	18
revised docs	247	151	0	449	215	3469	0	44	26
viewed lds	1805	392	840	2296	2027	21519	543	130	225
comment	39	0	0	0	0	491	0	0	6
delete	23	18	0	15	0	207	0	92	1

Table 6. Matrix form of Fig. 3

	login	create	create from d.	publish	revised docs	viewed lds	comment	delete	add editor	add viewer	del viewer	END
START	377	0	0	0	0	6	0	0	0	0	0	0
login	331	145	0	0	0	2498	0	0	0	0	0	52
create	0	1851	0	0	489	0	0	37	0	116	0	0
create from d.	8	0	9	11	98	0	0	15	0	0	0	0
publish	147	0	8	975	1035	943	0	0	0	184	0	20
revised docs	123	0	0	475	234	1986	0	0	0	471	3	13
viewed lds	1937	396	131	1718	1471	21938	806	88	8	434	16	276
comment	77	0	0	0	0	728	0	0	0	0	0	77
delete	0	28	0	0	0	147	0	335	0	0	0	4
add editor	1	0	0	0	0	0	0	0	11	20	0	0
add viewer	86	145	0	128	0	826	0	34	12	627	6	0
del viewer	0	0	0	0	5	24	0	3	0	0	19	1

Table 7. Matrix form of Fig. 4

	login	Create	Create from d.	Publish	Revised docs	Viewed lds	Comment	Delete	END
START	1	0	0	0	0	0	0	0	0
Login	0.09	0.03	0	0	0	0.85	0	0	0
Create	0.02	0.02	0.02	0.04	0.80	0.09	0	0.02	0
Create from dup.	0	0	0.03	0.05	0.77	0.10	0	0.04	0
Publish	0.06	0	0	0.25	0.42	0.25	0	0	0
Revised docs	0.06	0	0	0.03	0.04	0.83	0	0	0
Viewed lds	0.06	0.02	0.02	0.09	0.05	0.72	0.02	0	0
Comment	0.07	0	0	0	0	0.86	0.04	0	0
Delete	0.05	0.05	0.03	0.04	0.03	0.54	0	0.24	0

Table 8. Matrix form of Fig. 5

	login	create	create from d.	publish	revised docs	viewed lds	comment	delete	add editor	add viewer	del viewer	END
START	1	0	0	0	0	0	0	0	0	0	0	0
login	0.1	0.05	0	0	0	0.81	0	0	0	0	0	0.02
create	0	0.73	0	0	0.20	0	0	0	0	0.04	0	0
create from d.	0	0	0.05	0.05	0.66	0.10	0	0.08	0	0	0	0
publish	0.04	0	0	0.29	0.38	0.25	0	0	0	0.04	0	0
revised docs	0.05	0.05	0	0.03	0.06	0.63	0	0	0	0.16	0	0
viewed lds	0.07	0	0	0	0.04	0.75	0.01	0	0	0	0	0
comment	0.10	0	0	0	0	0.85	0.04	0	0	0	0	0
delete	0.05	0.05	0	0	0	0.26	0	0.60	0	0	0	0
add editor	0	0	0	0	0	0.69	0	0	0.31	0	0	0
add viewer	0.05	0	0	0.08	0.05	0.45	0	0	0	0.34	0	0
del viewer	0.03	0.06	0.05	0.09	0.06	0.36	0	0.03	0	0.03	0.29	0.05

References

1. Laurillard, D.: Teaching as a Design Science: Building Pedagogical Patterns for Learning and Technology. Routledge, Taylor & Francis Group, 7625 Empire Drive, Florence, KY 41042 (2012). http://www.routledge.com/books/details/9780415803878/++/
2. McKenney, S., Mor, Y.: Supporting teachers in data-informed educational design. Br. J. Educ. Technol. **46**(2), 265–279 (2015). https://doi.org/10.1111/bjet.12262
3. Mor, Y., Ferguson, R., Wasson, B.: Editorial: learning design, teacher inquiry into student learning and learning analytics: a call for action: learning design, tisl and learning analytics. Br. J. Educ. Technol. **46**, 221–229 (2015). https://doi.org/10.1111/bjet.12273
4. Laurillard, D., Kennedy, E., Charlton, P., Wild, J., Dimakopoulos, D.: Using technology to develop teachers as designers of TEL: evaluating the learning designer. Br. J. Educ. Technol. **49**(6), 1044–1058 (2018). https://doi.org/10.1111/bjet.12697
5. Culver, J.: The design of cloudworks: applying social networking practice to foster the exchange of learning and teaching ideas and designs. Comput. Educ. **54**(3), 679–692 (2010). https://doi.org/10.1016/j.compedu.2009.09.013

6. Dalziel, J.R.: Implementing learning design : the learning activity management system (LAMS) (2003)
7. Rodríguez-Triana, M.J., Prieto, L.P., Ley, T., de Jong, T., Gillet, D.: Social practices in teacher knowledge creation and innovation adoption: a large-scale study in an online instructional design community for inquiry learning. Int. J. Comput.-Support. Collab. Learn. **15**(4), 445–467 (2020). https://doi.org/10.1007/s11412-020-09331-5
8. Hernández-Leo, D., et al.: An integrated environment for learning design. Front. ICT **5**, 9 (2018). https://doi.org/10.3389/fict.2018.00009
9. Bennett, S., Lockyer, L., Agostinho, S.: Towards sustainable technology-enhanced innovation in higher education: advancing learning design by understanding and supporting teacher design practice. Br. J. Educ. Technol. **49**(6), 1014–1026 (2018). https://doi.org/10.1111/bjet.12683, https://bera-journals.onlinelibrary.wiley.com/doi/abs/10.1111/bjet.12683
10. Hernández-Leo, D., Martinez-Maldonado, R., Pardo, A., Muñoz-Cristóbal, J.A., Rodríguez-Triana, M.J.: Analytics for learning design: a layered framework and tools. Br. J. Edu. Technol. **50**(1), 139–152 (2019)
11. Recker, M., Yuan, M., Ye, L.: Crowdteaching: supporting teaching as designing in collective intelligence communities. Int. Rev. Res. Open Distance Learn. **15**, 138–160 (2014). https://doi.org/10.19173/irrodl.v15i4.1785
12. Veletsianos, G., Shepherdson, P.: A systematic analysis and synthesis of the empirical MOOC literature published in 2013–2015. Int. Rev. Res. Open and Distributed Learn. **17**(2), 198–221 (2016). https://doi.org/10.19173/irrodl.v17i2.2448
13. Siemens, G., Gasevic, D.: Guest editorial-learning and knowledge analytics. J. Educ. Technol. Soc. **15**(3), 1–2 (2012)
14. Clow, D.: An overview of learning analytics. Teach. High. Educ. **18**(6), 683–695 (2013). https://doi.org/10.1080/13562517.2013.827653
15. Van Der Aalst, W.: Process mining: overview and opportunities. ACM Trans. Manag. Inf. Syst. **3**(2), 7:1-7:17 (2012). https://doi.org/10.1145/2229156.2229157
16. Maldonado, J., Pérez-Sanagustín, M.: Analysis of students' self-regulatory strategies in MOOCS and their impact on academic performance. Ph.D. thesis (2020)
17. Juhaňák, L., Zounek, J., Rohlíková, L.: Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. Comput. Hum. Behav. **92**, 496–506 (2017). https://doi.org/10.1016/j.chb.2017.12.015
18. Maldonado, J., Pérez-Sanagustín, M., Kizilcec, R., Morales, N., Munoz-Gama, J.: Mining theory-based patterns from big data: identifying self-regulated learning strategies in massive open online courses. Comput. Hum. Behav. **80**, 179–196 (2017). https://doi.org/10.1016/j.chb.2017.11.011
19. Gutiérrez-Páez, N.F., Santos, P., Hernández-Leo, D., Carrió, M.: Designing a pre-service teacher community platform: a focus on participants' motivations. In: De Laet, T., Klemke, R., Alario-Hoyos, C., Hilliger, I., Ortega-Arranz, A. (eds.) EC-TEL 2021. LNCS, vol. 12884, pp. 352–357. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86436-1_34
20. Hur, J.W., Brush, T.A.: Teacher participation in online communities. J. Res. Technol. Educ. **41**(3), 279–303 (2009). https://doi.org/10.1080/15391523.2009.10782532
21. Hew, K.F., Hara, N.: Empirical study of motivators and barriers of teacher online knowledge sharing. Educ. Technol. Res. Dev. **55**(6), 573 (2007)
22. Denzin, N.K., Lincoln, Y.S.: The SAGE Handbook of Qualitative Research. SAGE (2011)

23. Mohammadi, F., Abrizah, A., Nazari, M., Attaran, M.: What motivates high school teachers to use web-based learning resources for classroom instruction? An exploratory case study in an Iranian smart school. Comput. Hum. Behav. **51**, 373–381 (2015). https://doi.org/10.1016/j.chb.2015.05.016, http://www.sciencedirect.com/science/article/pii/S0747563215003878
24. Michos, K., Davinia, H.L.: Supporting awareness in communities of learning design practice. Comput. Hum. Behav. **85**, 255–270 (2018). https://doi.org/10.1016/j.chb.2018.04.008
25. Hernández-Leo, D., Romeo, L., Carralero, M.A., Chacón, J., Carrió, M., Moreno, P., Blat, J.: LdShake: learning design solutions sharing and co-edition. Comput. Educ. **57**(4), 2249–2260 (2011). https://doi.org/10.1016/j.compedu.2011.06.016, https://www.sciencedirect.com/science/article/pii/S036013151100145X
26. van der Aalst, W.: Process Mining: Data Science in Action, 2nd edn. Springer, Cham (2016). https://doi.org/10.1007/978-3-662-49851-4
27. Saint, J., Fan, Y., Singh, S., Gasevic, D., Pardo, A.: Using process mining to analyse self-regulated learning: a systematic analysis of four algorithms, pp. 333–343 (2021). https://doi.org/10.1145/3448139.3448171
28. Weijters, A., Aalst, W., Medeiros, A.: Process mining with the heuristics miner-algorithm, vol. 166 (2006)
29. Gatta, R., et al.: pMineR: an innovative R library for performing process mining in medicine. In: ten Teije, A., Popow, C., Holmes, J.H., Sacchi, L. (eds.) AIME 2017. LNCS (LNAI), vol. 10259, pp. 351–355. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59758-4_42
30. Doberstein, D., Hecking, T., Hoppe, H.U.: Using sequence analysis to characterize the efficiency of small groups in large online courses. In: Proceedings of the 26th International Conference on Computers in Education. Philippines: Asia-Pacific Society for Computers in Education (2018)

Collaborative Community Knowledge Building with Personalized Question Recommendations

Alexander Tobias Neumann, Simon Breuer(✉), and Ralf Klamma(✉)

RWTH Aachen University, Aachen, Germany
{neumann,breuer,klamma}@dbis.rwth-aachen.de

Abstract. Inquiry-based learning focuses on asking questions and finding answers through investigation while engaging students in active learning. Decentralized knowledge building in the community can benefit from applying the approach to modern web technologies and collaboration. Distributed Noracle is a decentralized, question-based tool for building community knowledge in collaboration among participants that helps learners understand and reflect on individual learning and community processes. However, learners may encounter vast and complex space graphs, which leads to difficulties pursuing their goal of reflection and understanding the knowledge of the community. Community knowledge building is essential as it allows community members to share their knowledge and expertise. We extended the Distributed Noracle application with a recommender system that can suggest relevant questions and spaces to community members to address this problem. The recommendations can be accessed within the Distributed Noracle application or through our developed chatbot for community chat environments like Rocket.Chat or Slack. Our evaluation shows that our recommender system achieves the intended purpose and helps learners to find exciting questions in vast question spaces in a short amount of time.

Keywords: Community of practice · Knowledge building · Learning infrastructures · Recommender system · Chatbots

1 Introduction

In recent years, inquiry-based learning has become more and more popular. While in conventional and structured learning environments, people should mostly learn solid facts and answers to given questions. With inquiry-based learning as a pedagogical method, they can get a deeper understanding by asking questions and thinking of related concepts independently. This can also improve the critical thinking of the individual. Today, we know that students at all levels of education can successfully experience and develop deeper level thinking skills through scientific inquiry [18]. However, inquiry-based learning has not become an established practice in science education yet [10]. Even when it is considered an essential and valuable approach, teachers and instructors are unfamiliar

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 190–204, 2022.
https://doi.org/10.1007/978-3-031-20218-6_13

with it. Most of the time, people learn in so-called Communities of Practice (CoPs). While these communities do not have to be tied to a central organization or infrastructure, they have a common understanding of a profession or area. To support inquiry-based learning in CoPs, de Lange et al. [6] developed *Distributed Noracle*[1], a decentralized question-based dialog system. In this system, CoPs can create a space graph and start with a question or statement of curiosity, sometimes called the "wonder moment" [24]. Each member has the chance to expand the initial question with additional questions and mark questions as related. Additionally, the system can overview ignorance over a specific topic in a CoP by analyzing the uprising questions. Users can subscribe to different rooms where they can cooperate with the CoP. With multiple rooms, it's easy to lose track. The *Noracle Bot* helps users find their way around and navigate the rooms, providing awareness with information about recent activity in the room. A recommender system could help users by suggesting interesting questions they might not realize or overlook in the subscribed spaces to increase engagement. Thus, the system could help to overcome cognitive overload when users encounter vast and complex space graphs [19]. In this paper, we address the following research questions:
RQ1: What data sources create suitable recommendations for CoP in *Distributed Noracle*?
RQ2: How can we derive recommendations from such distributed architectures? Therefore we present a recommender system capable of suggesting relevant questions and spaces for a user inside the *Distributed Noracle* application.

The rest of the paper is structured as follows: The following section (Sect. 2) describes related work where we analyze different typical filtering approaches. After that, we describe the formal description of the retrieving process of recommendations (Sect. 3) followed by the implementation including the updated *Noracle Bot* (Sect. 4). In (Sect. 5) we present our evaluation and conclude with a summary and outlook (Sect. 6).

2 Related Work

2.1 Inquiry-Based Learning and Question-Based Dialog

The target audience of the *Distributed Noracle* is the aforementioned CoPs. Three elements define a CoP: Domain, Community, and Practice [28]. People with the same interests come together to learn and exchange knowledge. Inquiry-based learning can be seen as hands-on activities where learners generate meaningful questions with curiosity to get closer to the overall problem or question [17]. Learning is stimulated by inquiry, which means asking questions or solving problems [23]. In a traditional learning setting, the teacher works as a director, who gives the question and partly the answers to the learner;

[1] https://distributed-noracle.github.io.

in an inquiry-based learning setting, the teacher plays the role of a facilitator [24]. However, there is a large chunk of traditional schools where students are not encouraged to ask questions. Abd-El-Khalick already mentioned several dimensions where inquiry-based learning can be a valuable support, such as mathematical, linguistic, cognitive, and metacognitive skills, or personal, social, cultural, and ethical domains [1]. In large-scale investigations, this was shown especially for Science and Maths [3,12]. Asking questions is an essential part of learning and research. It is one of the thinking processing skills which is structurally embedded in the thinking operation of critical thinking, creative thinking, and problem-solving [5]. During COVID-19, the implementation of digital learning opportunities such as quizzes was criticized [7]. Learners found them unsatisfactory because the proposed quizzes did not assess the method of thinking, analyzing, writing, or understanding a problem, resulting in significant stress. Unfortunately, students or learners often do not have the opportunity to ask questions in school or university because they do not want to draw attention to themselves, or the teachers cannot motivate them. A face-to-face setting or question-based dialog between students and experts like teachers can then help explore the gap or discrepancy in the student's knowledge. It is important to emphasize that this does not necessarily increase learners' academic performance, as the quality of implementation still plays a significant role here; rather, it can have a positive impact on attitudes toward science [3].

In this paper, we utilize the Web-based application *Distributed Noracle*, which supports the inquiry-based learning approach by using the idea of a question-based dialogue.

2.2 Related Question Recommender Systems

Recommender systems, which act in question recommendation, can be derived from question-answering recommender systems. Zou et al. present a question-based recommender system, which constructs and algorithmically chooses questions [32]. The recommendations are based on a model that infers the underlying user belief and preferences over items. The model is trained offline with a matrix factorization algorithm and online with a closed-form solution based on the user's answers. Liu and Hao present a recommender system that suggests questions based on the relationship between the user and the question [15]. The relationship between users and the words in the corpus is calculated with an incremental update algorithm which takes questions and the users who answered them as input. Zhao et al. present an approach to identify question paraphrases in a vast collection of questions fetched from Encarta logs [21]. The approach is divided into five steps: *Question Extraction, Question Type Classification, Question Partition, Question Paraphrase Identification, Template Extraction.* The extraction of questions is done with filtering of log entries, which should contain at least three words and starts with an *WH-word (who, what, when, where, why, and how).* The question type is classified using a question word and a Support Vector Machine (SVM) classifier based on the question type taxonomy from Li and Roth [14] into 50 different sets. Each set is further partitioned into clusters, indexed by a specific content word, which appears in the questions assigned to the cluster. After that, the question paraphrasing is done by building question pairs inside the clusters and computing multiple features like *Cosine Similarity, Named Entity Overlapping, or WordNet Synonyms*. Ahasanuzzaman et al. provide a solution for identifying duplicate questions on Stack Overflow[2] by using a discriminative classification model called *Dupe* [2]. A preprocessing technique is used to process the questions in the model, where stemming and removing stop words from the title- and body-post is applied. After that, similar features that Zhao et al. [21] used are computed for each question pair (*Cosine Similarity Value* or *WordNet Similarity*). In other domains, such as online shops, software agents elicit the users' preferences and interests to make product recommendations [29]. Newer approaches like neural graph networks or neural graph collaborative filtering bear large potential [8,27]. They used standard datasets such as Amazon Book, Epinions, or Gowalla with millions of entries for the work's studies. *Distributed Noracle*, on the other hand, has only tiny interaction data, as the application is self-managed within communities. Since *Distributed Noracle* does not provide sufficient data, deep learning approaches are not suitable since they are known to be data-hungry [30]. A novel way to use recommender systems is the utilization of chatbots. Chatbots can be a more natural and interactive interface for users to interact with recommender systems and provide awareness [16,22]. Cerezo et al. developed a chatbot capable of recommending software artifact experts of an open-source project [4]. The users expected a chatbot that could hold a complete conversation beyond the recommender system functionality in their evaluation. This is not an easy task, and the authors also refer to the *uncanney valley* effect should be considered. Nevertheless, the recommender system was well received. The chatbot tries to classify the user messages using a term frequency algorithm to understand the intents. Laban and Araujo compared user-initiated and system-initiated conversational recommender systems regarding anthropomorphism, risk, and control perception [11]. Their results show that users feel they have less control and perceive the system as riskier when making system-initiated recommendations than user-initiated recommendations. Therefore, we are also considering a chatbot interface for our recommender system and letting the bot be part of the community.

3 Question Recommendation

For the question recommendation process, we consider the interactions or the behavior of different users in the *Distributed Noracle* application. Whenever two or more users have a lot in common, the recommendations for one user should be influenced by similar users. If we consider the interactions in the *Distributed Noracle*, users can subscribe to spaces, ask questions, create relations between questions and give up- and down-votes to questions and relations. From these interactions, several similarities can be derived. For example, users who asked questions with the same topic voted up or down the same questions and relations

[2] https://stackoverflow.com.

or had many common neighboring questions (connected via a relation). These similarities follow the *item-based* approach since the properties are all derived from the items, in this case, questions, relations, and votes. The user profile does not contain any relevant data for computing similarities between users.
In recommending questions, the recommender system tries to find the question $q' \in Q$, which should be the most exciting or relevant question for the user $u \in U$ in the current context. It takes several aspects derived from the *Distributed Noracle* application into account. We know which user asked which question, created which relation between questions, and gave up and down-votes to them. The corresponding formula for q' is shown in Eq. 1.

$$q' = \arg\max_{q \in Q} \, r(u, q) \tag{1}$$

The utility function $r(u, q)$, Eq. 2, computes the relevance of question q for user u:

$$\begin{aligned} r(u,q) = \; & \mathbf{1}_{(Q_u \cup V_u)^\complement}(q) \\ & * \Big(w_1 * content(u,q) \\ & + w_2 * collab(u,q) \\ & + w_3 * \frac{upVotes(q)}{\sum_{i=1}^{N} upVotes(q_i)} \\ & - w_4 * \frac{downVotes(q)}{\sum_{i=1}^{N} downVotes(q_i)} \\ & + w_5 * \big|\big|upVotes(q)\big|\big| \\ & - w_6 * \big|\big|downVotes(q)\big|\big| \\ & + w_7 * \big|\big|time(q)\big|\big| \Big) \\ & * \frac{1}{\sum_{i=1}^{7} w_i} \end{aligned} \tag{2}$$

which computes the final score of the relevance for question q for user u with an indicator function $\mathbf{1}_{(Q_u \cup V_u)^\complement}(q)$ shown in Eq. 3 and seven weighted functions with $w_{i \, \in [1,7]} \; \in \; (0, \; \infty)$ which we will explain in the following.

$$\mathbf{1}_{(Q_u \cup V_u)^\complement}(q) = \begin{cases} 1, & \text{if } q \in (Q_u \cup V_u)^\complement \quad Q_u = \{q \mid u \text{ asked } q\}, \\ 0, & \text{otherwise} \qquad\qquad\quad\;\; V_u = \{q \mid u \text{ voted } q\} \end{cases} \tag{3}$$

The indicator function is 1, if q was not created by u which means that $q \notin Q_u$ and u did not vote q which means that $q \notin V_u$, which implies $q \in (Q_u \cup V_u)^\complement$. Otherwise, the function returns 0. In other words, we only want to suggest questions that the user has not yet seen.

Now, we describe the seven weighted functions. Equation 4 shows the feature for w_1 describes the similarity between the contents of the questions.

$$content(u,q) = \max \, cosineSimilarity(q, q_u), \quad \forall q_u \in Q_u \tag{4}$$

Therefore, the *cosineSimilarity* shown in Eq. 5 is defined as follows:

$$cosineSimilarity(q_1, q_2) = \frac{W_{q_1} * W_{q_2}}{|W_{q_1}| * |W_{q_2}|} \tag{5}$$

W_{q_i} defines the vector consisting of the relative term frequencies of each word inside question q_i.

The feature for w_2 shown in Eq. 6 describes the similarity between the votes using the *Jaccard index*.

$$collab(u, q) = \max \frac{|V_u \cap V_v|}{|V_u \cup V_v|}, \quad \forall v \in U \wedge u \neq v \tag{6}$$

The $upVotes(q_i)$ function shown in Eq. 7 denotes the number of absolute up-votes of question q_i.

$$upVotes(q_i) = \big|\{v \mid v \in V_u \wedge v \text{ is a up-vote}\}, \forall u \in U\big| \tag{7}$$

Analogously, the $downVotes(q_i)$ function shown in Eq. 8 denotes the number of absolute down-votes of question q_i.

$$downVotes(q_i) = \big|\{v \mid v \in V_u \wedge v \text{ is a down-vote}\}, \forall u \in U\big| \tag{8}$$

That means that the features for weights w_3 and w_4 compute the relative number of up- and down-votes of question q, which is the input for the utility function $r(u, q)$.

The features for weights w_5 and w_6 are the normalized values of the up- and down-votes of question q.

The last feature for weight w_7 is the $time(q)$ function which is a normalized measure that tells how much time has passed since the question q was created.

4 Implementation

As the core backend of the *Distributed Noracle* application is based on multiple microservices in one project, we extended it with the needed microservices, *NoracleRecommenderService, NoracleNormalizationService,* and *NoracleQuestionUtilityService*, that are needed to realize the recommender system. The first extension is the additional REST-API resource called *RecommenderResource*, which can be requested for recommendations based on the *Distributed Noracle* space and the user. When the *RecommenderResource* is requested by a client, the corresponding functions in the *NoracleRecommenderService* for computing the recommended questions are invoked. The *NoracleRecommenderService* then collects all necessary data for that. Once all necessary data is collected, all questions are sent to the *NoracleNormalizationService* to be normalized. The *NoracleNormalizationService* will then apply natural language processing techniques to the questions. Most techniques (lowercasing, removing punctuation, removing hyphens, expanding contractions, removing digits, removing stop words) are realized with simple replacing operations. For *Stemming* we used the Porter-Stemmer library from Apache Lucene[3]. The algorithm considers each letter in a word as a consonant or a vowel for the steps. Each word can be described as $(C)^*(VC)^m(V)^*$ where C is a consonant, V a vowel and $m \in \mathbb{N}_0$. The rules for removing endings from words are given in the form of:

$$(condition)S1 \rightarrow S2 \tag{9}$$

where for example, $(V)ING \rightarrow \{\}$ means that we cut the suffix ING from every word that contains a vowel and trivially ends with ING. This rule transforms the word "walking" into "walk". We can illustrate that with the following two simple questions, which are also used in one space of our evaluation: q_1 = "What are the concepts of agile software development?" and q_2 = "What concept do you follow in agile software developing?"

Without stemming, the result of the cosine similarity with a naive word embedding approach:

$$V_{q_1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, V_{q_2} = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix}, cosineSimilarity(V_{q_1}, V_{q_2}) = \frac{V_{q_1} * V_{q_2}}{|V_{q_1}| * |V_{q_2}|} \approx 0.35 \tag{10}$$

In this case, the cosine similarity is approx. 35%, although the two questions mean more or less the same. With stemming, the cosine similarity can be increased to 59% as: q_1' = "What ar the concept of agil softwar develop?" and q_2' = "What concept do you follow in agil softwar develop"

$$V_{q_1'} = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix}, V_{q_2'} = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix}, cosineSimilarity(V_{q_1'}, V_{q_2\prime}) = \frac{V_{q_1'} * V_{q_2\prime}}{|V_{q_1'}| * |V_{q_2\prime}|} \approx 0.59 \tag{11}$$

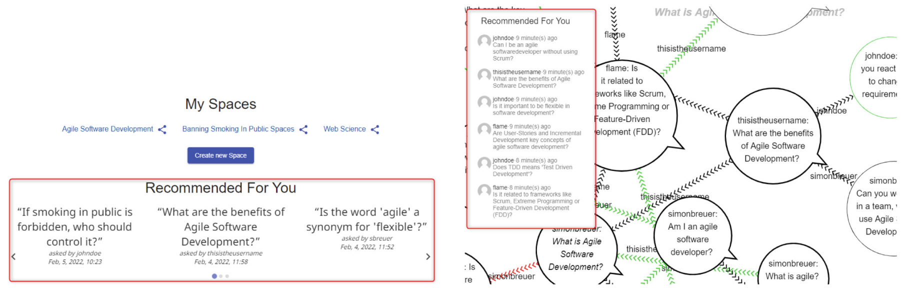
It is obvious, that the words "concepts", "development" and "developing" are reduced to their root, which explains the increased value of the cosine similarity. Nevertheless, words like "agil" and "softwar" do not make much sense anymore. This could result in questions where the meaning is heavily distorted. However, since no additional features, except the cosine similarity, on the stemmed words are computed, the distortion can be neglected. Up to know, the used Porter-Stemmer only works with English sentences or in our case questions. Thus, we are limited to the English language.

The last processing technique of the *NoracleNormalizationService* replaces words with synonyms. Therefore, we used WordNet[4] 3.0 and the java library JAWS[5], which is an API that is capable of retrieving data from a local WordNet database. With the integration of the WordNet database and the JAWS library in the *Distributed Noracle* backend, a list of synonyms is retrieved for each word

[3] https://lucene.apache.org.

[4] https://wordnet.princeton.edu.

[5] https://github.com/jaytaylor/jaws.



My Spaces

Recommended For You

"If smoking in public is forbidden, who should control it?" asked by johndoe Feb, 5, 2022, 10:23

"What are the benefits of Agile Software Development?" asked by thisistheusername Feb, 4, 2022, 11:58

"Is the word 'agile' a synonym for 'flexible'?" asked by sbreuer Feb, 4, 2022, 11:52

Recommended For You

johndoe 9 minute(s) ago
Can I be an agile softwaredeveloper without using Scrum?

thisistheusername 9 minute(s) ago
What are the benefits of Agile Software Development?

johndoe 9 minute(s) ago
Is it important to be flexible in software development?

flame 9 minute(s) ago
Are User-Stories and Incremental Development key concepts of agile software development?

johndoe 8 minute(s) ago
Does TDD means 'Test Driven Development'?

flame 8 minute(s) ago
Is it related to frameworks like Scrum, Extreme Programming or Feature-Driven Development (FDD)?

thisistheusername: What are the benefits of Agile Software Development?
simonbreuer: Am I an agile software developer?
simonbreuer: What is Agile Software Development?
simonbreuer: What is agile?
flame
thisistheusername
johndoe
simonbreuer

(a) Recommended questions inside the space overview.

(b) Recommended questions for a specific space.

Fig. 1. Integration of the recommendations within a *Distributed Noracle* space and in the overview.

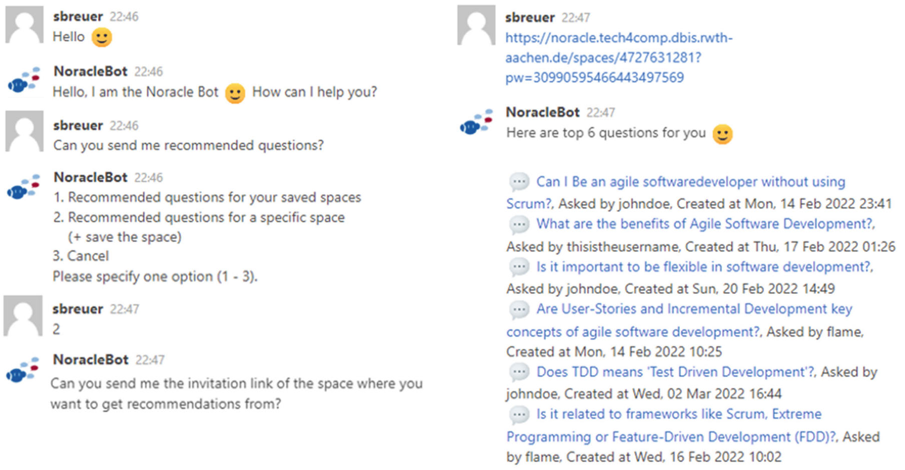
the question contains. For example, if we have the words "fright" or "dread" in the question, we replace it with the synonym "fear".

Once all potential questions have been normalized, the *NoracleRecommenderService* uses the *NoracleQuestionUtilityService* to compute a utility value for each question. The utility value is an indicator of the question's relevance to the corresponding user at the current time.

The next step was to adjust the *Distributed Noracle* frontend. It is realized as a web application where the user interface is the browser in which the user interacts. The first extension of the *Distributed Noracle* frontend is then made to the space overview page (see Fig. 1a). Three recommended questions are displayed at once, and the user can cycle through nine. Each slide contains one recommended question with its content, author, and creation time. The nine recommended questions have the highest utility value and are derived from all questions in the areas to which the user subscribes. If a user finds one of them interesting or valuable, he can click on it to be navigated to the corresponding space with the question selected. The related question and all neighboring questions are loaded and shown to the user. The second extension of the frontend displays six recommended questions within a space (see Fig. 1b). Again, clicking on one of the entries is possible to get to the corresponding question and its neighboring questions. The user gets the most relevant in one list with the computed personalized recommendations.

NoracleBot

As the *Noracle Bot's* purpose is to deliver recommended questions by guiding the user to them, we have a bot-driven conversation [4]. The only conversation intents the bot needs to understand are simple ones like *greeting*, *goodbye*, *help* etc. but trivially the most important one is the *recommendation* intent. If this intent is recognized, the bot requests the *RecommenderResource* and forwards it to the user. In order to define all intents for the bot, we used an Natural



sbreuer 22:46
Hello 🙂

NoracleBot 22:46
Hello, I am the Noracle Bot 🙂 How can I help you?

sbreuer 22:46
Can you send me recommended questions?

NoracleBot 22:46
1. Recommended questions for your saved spaces
2. Recommended questions for a specific space
(+ save the space)
3. Cancel
Please specify one option (1 - 3).

sbreuer 22:47
2

NoracleBot 22:47
Can you send me the invitation link of the space where you want to get recommendations from?

sbreuer 22:47
https://noracle.tech4comp.dbis.rwth-aachen.de/spaces/4727631281?pw=30990595466443497569

NoracleBot 22:47
Here are top 6 questions for you 🙂

Can I Be an agile softwaredeveloper without using Scrum?, Asked by johndoe, Created at Mon, 14 Feb 2022 23:41
What are the benefits of Agile Software Development?, Asked by thisistheusername, Created at Thu, 17 Feb 2022 01:26
Is it important to be flexible in software development?, Asked by johndoe, Created at Sun, 20 Feb 2022 14:49
Are User-Stories and Incremental Development key concepts of agile software development?, Asked by flame, Created at Mon, 14 Feb 2022 10:25
Does TDD means 'Test Driven Development'?, Asked by johndoe, Created at Wed, 02 Mar 2022 16:44
Is it related to frameworks like Scrum, Extreme Programming or Feature-Driven Development (FDD)?, Asked by flame, Created at Wed, 16 Feb 2022 10:02

Fig. 2. Conversation with the NoracleBot. User greets the bot, invites NoracleBot to the Space and receives recommendations.

Language Understanding (NLU) Model[6]. When choosing the messenger, we paid attention to the community's needs and chose Rocket.Chat, which uses the same single sign-on. The user has three options to interact with the *Noracle Bot*. (1) He can either request recommended questions for his saved spaces, (2) request recommended questions for a specific space, or (3) cancel the conversation. If the first option is specified, the *Noracle Bot* will look up the saved spaces of the user, triggers the computation of the recommendations, and sends them back to the user. For the second option, the bot will first ask for the invitation link of the space where the user wants to get recommendations (see Fig. 2). The *Noracle Bot* as an agent needs to be a participant of the space to gain read access to all questions, relations, votes, etc. and thus, to be able to request the *NoracleRecommenderService* for recommendations. With the bot's subscription to the space, he becomes a part of the corresponding community because he is equivalent to other participants. After sending the invitation link, the user only needs to wait for the top six recommended questions from the corresponding space. A considerable advantage of the *Noracle Bot* is that the user can retrieve the recommendations via a conversation. Without it, users could only login into the *Distributed Noracle*, search the specific space and watch out for the recommended questions. While in the chat interface, independent of the used device, the *Noracle Bot* is always accessible. In addition, the *Noracle Bot* helps to address the long-standing challenges of growing communities [20] by being a community member himself.

[6] https://rasa.com.

5 Evaluation

In our evaluation, we asked users from a CoP to use the developed recommender system to evaluate the usability and usefulness of the given recommendations We recruited a CoP with 17 members fitting the mentioned target audience for inquiry-based learning [1,3,12]. Therefore, we prepared three spaces with different topics of computer science and culture ("Agile Software Development", "Banning Smoking In Public Places", and "Web Science") with more than 20 questions/relations each. First, they should register or login into the *Distributed Noracle* application. After that, they should join the three different spaces and give an up- or down-vote to at least three questions per space. The users should then consider and test the recommended questions inside the space overview and the spaces themselves. When they tested the recommended questions in the frontend, they should also contact the *Noracle Bot* and find out how to get the recommendations from him. Ultimately, the participants are asked to fill out an online survey. The participants had to reconsider the six recommended questions of a space of choice. In the survey, they had to write down the number of questions (a subset of the six recommended questions) relevant or exciting to them. In addition to that, they had to write the number of questions in the space but not in the list of recommended questions. The survey consists of general demographic questions, questions regarding the recommended questions inside the frontend and inside the chat with the *Noracle Bot*, and free text questions belonging to the recommender system. In total, 12 participants identified as male, three as female, and two did not answer the question of their gender. Regarding the age, 16 participants were between 20 and 30 years old, six were between 31 and 40 years old, one was between 41 and 50, and one gave no answer.

5.1 Results

The created survey contains questions regarding the general impression of the system as well as the ten statements of the *System Usability Scale* and some questions to measure the success of the recommender system. The participants agreed that the recommended questions helped to get a quick overview of the most relevant questions. A few participants mentioned that the navigation when clicking on a recommended question saves time. The next step was to evaluate the usability of the recommender system with the *System Usability Scale*. It consists of 10 statements where the participants need to respond with one of five options: from *strongly disagree* to *strongly agree*.

Most of the users understand the usage and the idea of the recommender system. The results are shown in Table 1. In our case we have a *System Usability Scale* score of $S \approx 68$ which is the average of all *System Usability Scale* studies [26]. Most participants found the introduction of a chatbot very interesting and were curious when interacting with the bot.

The participants subsequently classified the questions between interesting and not interesting. It allowed us to create a confusion matrix and calculated recall, precision, and accuracy. The corresponding confusion matrix is shown in Table 2, where "positive" corresponds to the number of interesting or relevant

Table 1. Participants' evaluation of the statements about the System Usability Scale of the integrated recommender system on an ordinal scale from 1 $\widehat{=}$ "strongly disagree" to 5 $\widehat{=}$ "strongly agree" ($n = 17$).

#	Statement	$\bar{x}_i$	σ_i
Q1	I think that I would like to use the Recommender System frequently	2.88	0.9
Q2	I found the Recommender System unnecessarily complex	2.24	1.06
Q3	I thought the Recommender System was easy to use	3.88	0.76
Q4	I think that I would need the support of a technical person to be able to use the Recommender System	1.82	1.1
Q5	I found the various functions in this system were well integrated	3.71	0.89
Q6	I thought there was too much inconsistency in this Recommender System	2.13	0.85
Q7	I would imagine that most people would learn to use this system very quickly	3.83	0.96
Q8	I found the Recommender System very cumbersome to use	2.91	1.25
Q9	I felt very confident using the Recommender System	3.52	1.14
Q10	I needed to learn a lot of things before I could get going with this Recommender System	2.65	1.27

Table 2. Confusion matrix from participant rating of recommended questions.

		Predicted	
		Positive	Negative
Actual	Positive	$TP = 68$	$FN = 28$
	Negative	$FP = 74$	$TN = 261$

recommendations and "negative" corresponds to the not interesting or irrelevant questions. All spaces contained 23 questions. With 17 participants, we got 68 questions recommended "correctly". Further, we have 74 questions recommended "incorrectly" (FP) as well as 28 questions that were not recommended but relevant (or interesting) for the user. In total, 261 questions were not recommended "correctly". While the *Precision* (47.89%), which is the number of recommended questions, that are relevant for the user, was very low, the *Recall* (70.83%) and the *Accuracy* (81.64%) were relatively high.

$$Precision : P = \frac{TP}{TP + FP} = \frac{68}{68 + 74} \approx 47.89\% \quad (12)$$

$$Recall : R = \frac{TP}{TP + FN} = \frac{68}{68 + 28} \approx 70.83\% \quad (13)$$

$$Accuracy : A = \frac{TP + TN}{TP + TN + FP + FN} = \frac{68 + 261}{68 + 261 + 74 + 28} \approx 81.64\% \quad (14)$$

5.2 Discussion

In general, the results can be interpreted primarily as neutral to positive. The recommender system scored in features like ease of use, no complexity, and good integration. With a *System Usability Scale* score of 67.7 (on a scale from 1 to 100), the system is slightly below average but also leaves room for improvement. For example, questions they have already interacted with should be suggested less often. At the same time, one could increase the system's transparency by showing why a question is recommended. Further, we evaluated the performance properties precision (47.89%), recall (70.83%), and accuracy (81.64%) with the construction of a confusion matrix to get a first impression of how good the recommender system achieves its goal. Although some questions were not always relevant to the user, the recommendation system barely recommended questions that were not relevant. This is also related to our answer to our first research question. In our approach, we only considered a few features available to use with the *Distributed Noracle* application. We only used cosine similarity and vote similarity for the personalized recommendations, which could explain the low precision. However, it must be considered that it was not easy to decide if a recommended question was a good candidate for the users. The recall and accuracy are significantly higher, which shows that the performance of the recommender system is more positive in the corresponding tasks. With the information gained from this work, the promising approaches of Fan et al. and Wang et al. can be adapted to this approach in the future [8,27]. However, it is worth mentioning that Fan et al. valued the interaction graph and neglected side information on users [8]. In the future, datasets from Zhihu[7] or Stack Overflow can be used to evaluate the recommender system [13] further.

6 Conclusion and Future Work

In this paper, we constructed a recommender system for the *Distributed Noracle* application, which is capable of recommending personalized questions inside the frontend as well as inside a chat interface in Rocket.Chat with the *Noracle Bot*. These recommendations are a subset of all questions from the space to which the user has subscribed. For the construction, we analyzed the different typical filtering approaches where content-based and collaborative filtering seem suitable in the case of the *Distributed Noracle*. We derived different feature values from these filtering approaches, like the cosine and vote similarity, which are then used with simple question properties and corresponding weights in the final utility computation to rank the possible recommended questions (RQ1). We relied on the microservice architecture to access the corresponding data derived from the answer for RQ1 within the decentralized application (RQ2). The implementation as a microservice and the modular design allowed us to make parts of the system, like normalizations and stemming, reusable by other services. In the evaluation, we used the system usability scale to understand how well the system

[7] https://www.zhihu.com.

performs in the frontend and with many different characteristics of the *Noracle Bot*. During the evaluation process, it came out that the recommender system has a slightly positive impact on the learning success and the user experience, especially for the members of the CoP. Further, it offers added value for encountering the huge cognitive load by getting a fast overview of the most relevant questions when users have to deal with spaces with at least 25 questions. With the results of the evaluation and related work in question recommendation, the recommender system can be further extended, evaluated, and improved. Some participants pointed out that there is no transparency regarding the delivery of the recommended question. A typical example would be "Because you asked question x" when the cosine similarity is above a certain threshold which indicates similar questions as the user asked. Other simple hints would be "New" or "Trending" when a question was asked recently or has many up-votes.

While the current recommender system uses a feature weighting scheme, the next step could be introducing an approach in machine learning and including semantics. The corresponding model could be built upon the used features as input and a dataset like Zhihu as training data [13]. Then the weights would lie in the model to learn concerning the used training data from the set. The model could then be compared with other state-of-the-art recommender system models in the area of Community Question Answering (CQA) [9,25,31]. In addition, the Normalized Discounted Cumulative Gain (NDCG) would be a suitable measurement to evaluate the performance of the model because it considers the ranking on the final recommendations.

The recommender system can be considered a valuable tool for the *Distributed Noracle* application.

Acknowledgements. The research leading to these results has received funding from the German Federal Ministry of Education and Research (BMBF) through the project tech4comp (grant no. 16DHB2110).

References

1. Abd-El-Khalick, F., et al.: Inquiry in science education: international perspectives. Sci. Educ. **88**(3), 397–419 (2004)
2. Ahasanuzzaman, M., Asaduzzaman, M., Roy, C.K., Schneider, K.A.: Mining Duplicate Questions in Stack Overflow. In: Kim, M., Robbes, R., Bird, C. (eds.) 13th Working Conference on Mining Software Repositories - MSR 2016, pp. 402–412. IEEE, Piscataway, NJ (2016)
3. Cairns, D., Areepattamannil, S.: Exploring the relations of inquiry-based teaching to science achievement and dispositions in 54 countries. Res. Sci. Educ. **49**(1), 1–23 (2019)
4. Cerezo, J., Kubelka, J., Robbes, R., Bergel, A.: Building an expert recommender chatbot. In: 2019 IEEE/ACM 1st International Workshop on Bots in Software Engineering (BotSE), pp. 59–63. IEEE, [Place of publication not identified] (2019)
5. Chin, C., Osborne, J.: Students' questions: a potential resource for teaching and learning science. Stud. Sci. Educ. **44**(1), 1–39 (2008)

6. de Lange, P., Goschlberger, B., Farrell, T., Neumann, A.T., Klamma, R.: Decentralized learning infrastructures for community knowledge building. IEEE Trans. Learn. Technol. **1** (2020)
7. Loubière, K., et al.: Attempts, successes, and failures of distance learning in the time of COVID-19. J. Chem. Educ. **97**(9), 2448–2457 (2020)
8. Fan, W., et al.: Graph neural networks for social recommendation. In: Liu, L., White, R. (eds.) The World Wide Web Conference on - WWW 2019, pp. 417–426. ACM Press, New York, New York, USA (2019)
9. Fang, H., Wu, F., Zhao, Z., Duan, X., Ester, M., Zhuang, Y.: Community-based question answering via heterogeneous social network learning, pp. 122–128 (2016)
10. Hofer, E., Lembens, A.: Putting inquiry-based learning into practice: how teachers changed their beliefs and attitudes through a professional development program. Chem. Teach. Int. **1**(2) (2019)
11. Laban, G., Araujo, T.: The effect of personalization techniques in users' perceptions of conversational recommender systems. In: Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents, pp. 1–3. ACM, New York, NY, USA (2020)
12. Lazonder, A.W., Harmsen, R.: Meta-analysis of inquiry-based learning. Rev. Educ. Res. **86**(3), 681–718 (2016)
13. Li, N., Guo, B., Liu, Y., Yao, L., Liu, J., Yu, Z.: AskMe: joint individual-level and community-level behavior interaction for question recommendation. World Wide Web **25**(1), 49–72 (2022)
14. Li, X., Roth, D.: Learning question classifiers. In: Proceedings of the 19th International Conference on Computational Linguistics-volume 1, pp. 1–7 (2002)
15. Liu, G., Hao, T.: User-based question recommendation for question answering system. Int. J. Inf. Educ. Technol. 243–246 (2012)
16. Mahmood, T., Ricci, F.: Improving recommender systems with adaptive conversational strategies. In: Cattuto, C. (ed.) Proceedings of the 20th ACM conference on Hypertext and hypermedia, p. 73. ACM Conferences, ACM, New York, NY (2009)
17. Minner, D.D., Levy, A.J., Century, J.: Inquiry-based science instruction-what is it and does it matter? Results from a research synthesis years 1984 to 2002. J. Res. Sci. Teach. **47**(4), 474–496 (2010)
18. National Research Council - Committee on Human Factors: National Science Education Standards. The National Academic Press (1996)
19. Paas, F., van Merriënboer, J.J.G.: Cognitive-load theory: methods to manage working memory load in the learning of complex tasks. Curr. Dir. Psychol. Sci. **29**(4), 394–398 (2020)
20. Seering, J., Luria, M., Kaufman, G., Hammer, J.: Beyond dyadic interactions: considering chatbots as community members. In: Brewster, S., Fitzpatrick, G., Cox, A., Kostakos, V. (eds.) Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–13. ACM, New York, NY, USA (2019)
21. Zhao, S., Zhou, M., Liu, T.: Learning question paraphrases for QA from Encarta logs. In: Manuela M. Veloso (ed.) IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, 6–12 January 2007, pp. 1795–1801 (2007)
22. Shum, H.Y., He, X.D., Li, D.: Challenges and opportunities with social chatbots. Front. Inf. Technol. Electr. Eng. **19**(1), 10–26 (2018)
23. Spronken-Smith, R., Walker, R.: Can inquiry-based learning strengthen the links between teaching and disciplinary research? Stud. High. Educ. **35**(6), 723–740 (2010). https://doi.org/10.1631/FITEE.1700826

24. Suarez, A., Ternier, S., Kalz, M., Specht, M.: GPIM: Google glassware for inquiry-based learning. In: Rensing, C., de Freitas, S., Ley, T., Muñoz-Merino, P.J. (eds.) EC-TEL 2014. LNCS, vol. 8719, pp. 530–533. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11200-8_58
25. Tu, H., Wen, J., Sun, A., Wang, X.: Joint implicit and explicit neural networks for question recommendation in CQA services. IEEE Access **6**, 73081–73092 (2018)
26. Vaziri, D.D., et al.: Exploring user experience and technology acceptance for a fall prevention system: results from a randomized clinical trial and a living lab. Eur. Rev. Aging Phys. Act. Official J. Eur. Group Res. Elderly Phys. Activity **13**, 6 (2016)
27. Wang, X., He, X., Wang, M., Feng, F., Chua, T.S.: Neural graph collaborative filtering. In: Piwowarski, B., Chevalier, M., Gaussier, E., Maarek, Y., Nie, J.Y., Scholer, F. (eds.) Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 165–174. ACM, New York, NY, USA (2019)
28. Wenger, E.: Communities of Practice: Learning, Meaning, and Identity. Learning in doing, Cambridge University Press, Cambridge, UK (1998)
29. Xiao, B., Benbasat, I.: E-commerce product recommendation agents: use, characteristics, and impact. MIS Q. **31**(1), 137–209 (2007)
30. Zhang, S., Yao, L., Sun, A., Tay, Y.: Deep learning based recommender system. ACM Comput. Surv. **52**(1), 1–38 (2020)
31. Zhao, Z., Yang, Q., Cai, D., Yueting, H., Zhuang, W.: Expert finding for community-based question answering via ranking metric network learning. In: Kambhampati, S. (ed.) Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, pp. 3000–3006. AAAI Press/International Joint Conferences on Artificial Intelligence, Palo Alto, California (2016)
32. Zou, J., Chen, Y., Kanoulas, E.: Towards question-based recommender systems. In: Huang, J. (ed.) Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 881–890. ACM Digital Library, Association for Computing Machinery, New York, NY, United States (2020)

Evaluating an Adaptive Intervention in Collaboration Scripts Deconstructing Body Image Narratives in a Social Media Educational Platform

René Lobo-Quintero(✉), Emily Theophilou, Roberto Sánchez-Reina, and Davinia Hernández-Leo

TIDE, ICT Department, Universitat Pompeu Fabra, Barcelona, Spain
{renealejandro.lobo,emily.theophilou,roberto.sanchez,
davinia.hernandez-leo}@upf.edu

Abstract. Social Media is an important disseminator of body image representations and the body cult. The growing popularity of social media among children and adolescents makes minors a vulnerable group to the internalization of body ideals and stereotypes. Developing educational interventions that provide adolescents with skills to better understand the body image in social media is therefore necessary to counteract the effects of deceitful representations and discourse. This paper evaluates an adaptive educational intervention to define the suitable approach to teach adolescents about body image and stereotyping in social media. In particular, the paper examines and compares three approaches to identify the dominant body image stereotype in students' social media: The self-reported methods, the analysis of social preferences, and the use of xAPI to track users' behavior. Results showed that the use of xAPI combined with self-reported answers can provide better input from adolescents' preferences. Moreover, it allows the automatic distribution of suitable counter-narratives to students participating in computer-supported collaborative learning activities embedded in an educational social media platform.

Keywords: Social media · Digital skills · Self-protection skills · CSCL scripts · Counter-narratives

1 Introduction

Social media platforms are important mediators in the construction of youngsters' body image [1]. While liking, tagging, and sharing social media posts, youngsters are exposed to different types of content, including representations that normalize the body ideals and stereotypes [2]. Image exposure and content interaction can reinforce the negative attitudes and perceptions that affect youngsters' body image [3–5], which can lead teenagers to experience body image dissatisfaction [6]. While some experts argue stronger regulations and even prohibition may be a solution [7], the media education of adolescents towards the uses and implications of social media may also be a fair measure [8, 9].

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 205–217, 2022.
https://doi.org/10.1007/978-3-031-20218-6_14

Developing ad hoc interventions can help children protect themselves, especially when raising awareness for the well-being is spoiled by both psychological and cultural factors [10]. One exemplary intervention is the one promoted by the use of narrative scripts (NS) [11] a combination of CSCL and story-telling techniques. As a pedagogical approach, NSs has been defined as an opportunity to raise awareness by situating young users into social media tailored risk scenarios and empowering their critical thinking attitudes.

NSs seek to provide students with learning material based on their own social media experiences. Nevertheless, to meet students' experience, NSs must cope with students' social media reality; for instance, preferences, patterns of interaction, time of exposure, etc. Adapting learning content to the user's behavior is not a new topic in research as adaptive learning systems have been on the forefront for some years now. Adaptive learning technologies (ALT) collect data based on the students' performance and behavior, and provide them with personalized learning materials and feedback. Integrating ALT to body image literacy can contribute to a better understanding of body image discourse and stereotypes by providing students with learning material based on the stereotypes they are exposed to.

In the topic of body image, ALT has yet to explore techniques to identify students' body image stereotypes. As the use of social media among adolescents is constantly growing and it can pose a threat to their body image perception, we find it important to explore this line of work. Therefore, the aim of this study is to explore different techniques to determine the students' dominant stereotypes in the topic of body image in order to facilitate the creation of CSCL activities.

1.1 Adaptive Learning Technology and CSCL Scripts for a Social Media Literacy Intervention

ALT is considered an emerging educational technological innovation [12]. It has pedagogical benefits, including acceleration, remediation, metacognition, mastery-based learning, immediate feedback and interactive learning [13] The Horizon Report (2018) explains that adaptive learning occurs when digital tools and systems are used to create individual learning paths for students based on their strengths, weakness and peace of learning however some adaptive learning systems include profile information from other sources [14].

One of the key components of Adaptive Learning is the learner model. The learner model refers to a student's representation as observed by an intelligent tutor. It collects data related to the student's behavior and performance within a virtual platform and reasons about adjusting feedback [15]. Besides collecting data related to the learning progress (activities, material, performance) ALT is also capable of identifying students' personality traits based on their behavior and performance. Research on personality traits and ALT saw the implementation of systems capable of understanding students cognitive and affective states [16]. In the topic of body image, ALT has yet to explore techniques to identify students' dominant stereotypes. As the use of social media among adolescents is constantly growing and it can pose a threat to their body image perception, we find it important to explore this line of work.

CSCL scripts structure a collaborative learning flow (group formation, sequence of tasks, role rotation, etc.) to facilitate the triggering of desired social interactions leading to fruitful learning by defining the task structure, the time structure, and the social structure of the collaborative activities [17]. Some of its applications include organizing collaboration by grouping students with different opinions to work on a task (ArgueGraph [18]), distributing information between participants for later discussion and explanation (jigsaw CLFP) and achieving consensus between participants by progressive knowledge building (Pyramid CLFP). ALT has been integrated in the design of CSCL systems to maximize the user-tailored support provided to group learners, focusing both on improved domain learning and the development of collaboration skills. The potential benefits of ALT in CSCL are high, by analyzing the user traits and characteristics a better group formation and knowledge distribution can be achieved or even adapting the activity to the group collaboration in real-time [19].

1.2 Exploring Approaches for the Detection of Body Image Dominant Stereotypes

Acquiring information about the learner model is often related with machine learning techniques, however for the identification of personality traits non-machine learning approaches are also used. A few such approaches have seen the use of pattern recognition techniques, self-labeling, questionnaires, and in some cases a combination of the two approaches has been applied [20–22]. In this section, we will identify existing approaches that are currently used in the topic of body image to identify the existence of stereotypes primarily outside the area of adaptive learning.

The use of questionnaires appears to be a preferred method in media literacy studies to identify stereotypes in social media. The study by Verrastro et al. [23] saw the implementation of a set of questionnaires on the topic of Instagram use and body-related scales to study the relationship between the use of Instagram, the internalization of beauty standards, the social pressure to adhere to them, and anxiety towards body image. Other approaches saw the implementation of open and self-reported answers to understand potential stereotypes. One such approach is the one described by Niemann et al. [24]. They performed a cluster analysis on a set of open-answers given by students in a survey about adjectives associated with demographic groups. According to them "findings indicate that open-ended responses, although laborious to organize, can be successfully employed for stereotype research". Despite being a laborious work, the use of self-reported answers has the potential to provide insights about users' preferences in social media as it receives a direct input from the users.

A more complex approach saw a content analysis to understand stereotypes that appeared within Instagram profiles. Butkowski et al. [25] studied how the stereotypical gender display was presented in young adult women's Instagram selfies, due to the complexity of the task they performed a content analysis using a manual coding scheme to classify the different variables (type of pose, expressed emotions, amount of body display) and a quantitative analysis of the feedback of the studied selfies (number of likes and comments).

As the use of questionnaires and content analysis will only provide us with input based on the result of a one-time analysis, we further explored possibilities of implementing approaches that would collect information based on the user's behavior. We identified a previous work that has attempted to extract data from social networks using web scraping tools and store it in a Learning Record Storage using the experience API (xAPI) data structure enabling teachers to create different learning activities based on the student's behavior [26].

The xAPI is a community-driven specification for learning technology. It was born from applying the Activity Streams concept to e-learning [27]. It defines both data and communication models to track user activities within learning software applications and has been used in different learning scenarios, such as serious games [28], online learning [29] and self-regulated learning [30]. In xAPI each event is captured as a statement, that is formed as a sentence with an actor (student token id), verb (action performed), activity (object) and context (time, session id, environment), that is stored chronologically in JSON format. The flexibility of xAPI allows the inclusion of different types of variables in the statements and the wide vocabulary of verbs already contains the most used social networks interactions allowing us to successfully track all the student's actions inside the educational social media platform.

The use of an appropriate method to identify students body image stereotypes will allow educators to provide students with material to counter possible toxic body ideals that arise. One of the strongest components of implementing ALT for raising awareness of body image stereotypes is the implementation of counter-narratives [31]. The implementation of counter-narratives within a social media can take advantage of different factors such as the allocation of participants into different scenarios based on their content preferences and interactions. As participants get involved in different interaction patterns while using social media, counter-narratives can be adapted to the personal interests and learning needs of young users.

Considering to date no study has considered this approach to examine students' interaction with body image content, the aim of this work is to explore different techniques to categorize learners based on the students' dominant stereotypes in the topic of body image. To do that, we pursued these specific objectives:

O1. Analyze teens' social media use and content preferences.
O2. Identify teens' exposure to body image content.
O3. Observe and analyze teens' SM interactions/online behavior within the designed educational platform.

2 Method

2.1 Study Design and Sample

This study explored three approaches to determine the dominant stereotype that students may have when exploring social media. The data for these approaches were gathered during two sessions (4 h) of digital literacy workshops carried out in three schools. During these sessions, 186 students (n = 186; 87 male, 88 female, and 11 undefined; Ages 13 to 16, mean age = 13.9, SD = 0,74) answered a questionnaire, registered to an educational social media platform and accessed a narrative script that covered different topics of threats and dangers that exist within social media. The workshops took place during school hours and therefore the students were placed in their assigned classrooms.

For this study, an initial focus was made on beauty and body stereotypes that can be created by influencers on Instagram. A narrative script has been designed to expose the reality behind the curated content of influencers falling under the categories of beauty and fitness.

The categories of beauty and fitness have been chosen as they target the topic of idealized body image that immensely exists in social media platforms. We have classified as beauty influencers, accounts that tend to share content related to make-up tutorials, high fashion, and modeling. The category of fitness includes accounts that focus on idealized body image, muscles, workouts and diets.

The third category of neutral influencers has also been considered to act as a gateway in case a student does not show a particular interest in the aforementioned categories. The neutral influencers refer to influencers that do not promote a specific beauty or body stereotype and are more focused on music, games, or traveling.

2.2 Measurement and Instruments

The research design was embedded in two sessions of the digital literacy workshops, participants were requested to perform different tasks and answered a questionnaire expressly designed for this study. Then they registered and navigated in a simulated educational social network, that was preloaded with profiles and photos that had the narratives of fitness, beauty, and neutral, but were not different from a regular profile to not bias the student's perception. In total 16 female and male predefined influencer profiles were created; 4 fitness profiles, 4 beauty profiles, and 8 neutral profiles.

Each of the interactions with the preloaded content was stored using the xAPI structure using the verbs liked, commented, opened (a specific profile), viewed (a specific photo) followed (a specific profile), and comment (a photo). In addition, each predefined profile was assigned a category that was captured by the xAPI.

The preloaded content was displayed within the same timeline as the student's content. Taking into consideration that some curiosity could rise as to who this person is, we considered assigning data weights to each xAPI verb to minimize curiosity from hindering the user's real preferences. The data weights were assigned to reflect the level of interest in each action. For the above verbs the following weights were considered; user viewed an image or video, 1 point, user liked or commented on a photo or video, 2 points, a user opened a profile page, 4 points, and finally, if a user followed a profile, 6 points were awarded.

For the purpose of this study, we focused on three variables:

- Social Media Use: measured from the questionnaire by asking the students about the top influencers that they followed.
- Body Image Source of Information: measured from the questionnaire through a list of topics and information sources.
- Online behavior: this variable was observed through the data captured using xAPI when students interacted with the predefined influencer profiles.

2.3 Procedure

At the beginning of the first workshop session, students answered a set of questions where they selected which sources of information, they use to get information about the topics of fitness and beauty. The questions included which sources (social media, friends and family, advertisement, TV and other communication media, experts on the topic) they use to keep up with topics such as fashion, personal appearance, weight loss, exercise, nutrition, and muscle gains. The number of sources they consulted expressed their interest in these topics, they could also select that they didn't use any source of information, meaning that they were not interested in the topic. Then, students were asked to indicate the 3 top Instagram profiles (accounts with more than 10000 followers) they follow daily.

During the two sessions of the workshop, the students were given approximately 20 min, each session, for free-roaming on the educational platform, where they had the opportunity to publish their content and interact with the content already published in the platform and by their classmates. Following the data collection, an algorithm was developed that would allocate a role (beauty, fitness, or neutral) to each student based on the footprint they left behind them.

3 Results

3.1 Students' Social Media Use

The self-reported answers show that students are interested in both topics, in total the sum of the sources used by students to get information about fitness was 577 and 549 about beauty (Fig. 1). However, only 27 students can be considered to be highly interested in

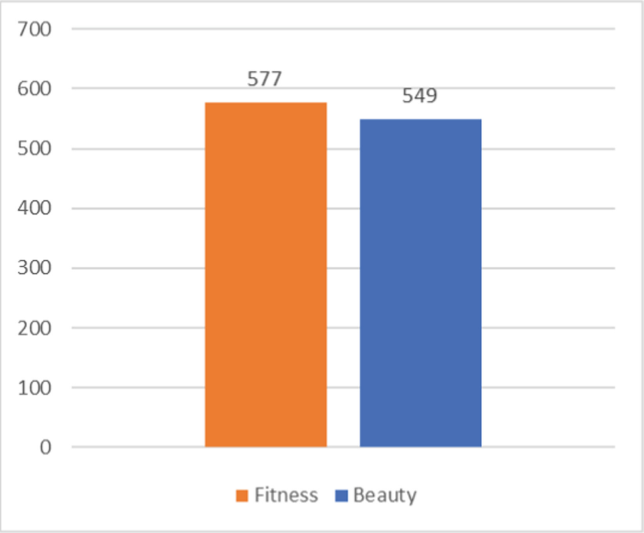


Fig. 1. Total sum of the information sources by each category

the topics (outside of the third quartile), with a score of 5 information sources or more of the maximum of 15 (Fig. 2).

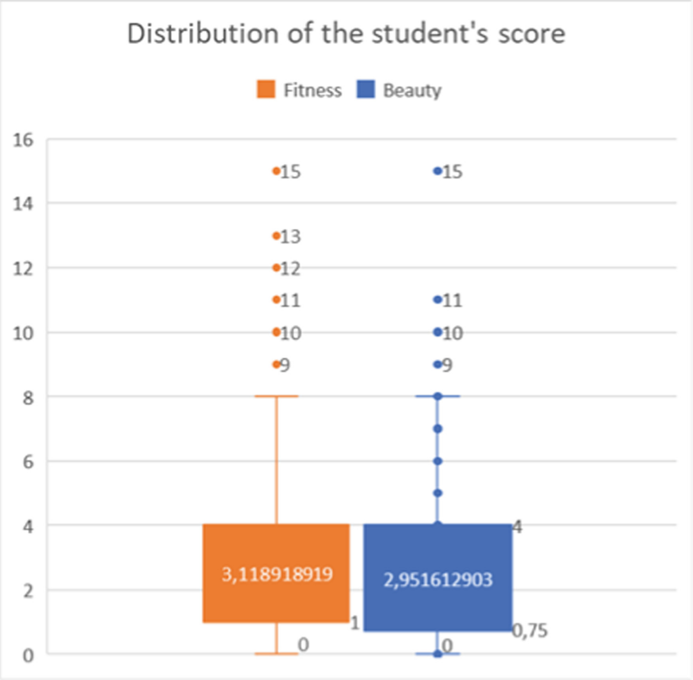


Fig. 2. Distribution of the scores of the information sources by each category

142 students of the 186 total students selected at least one source of information. The distribution of their scores shows different levels of interest in the topics, as can be seen in Fig. 3.

3.2 Body Image Content and Influencer Preferences

A classification of the category of each influencer was performed manually by research team members. In total, 291 influencers were reported with 181 being unique as some of the most famous influencers were reported more than one time. This classification by categories allowed us to estimate the general narrative of the group showing similar results to the ones of the first approach. However, as it can be seen in Fig. 4, students expressed an interest in influencers of the neutral type, that are related to topics like gaming, traveling, or entertainment.

3.3 Students' Online Behavior

During the study, the xAPI registered 723 interactions created by 102 students. To limit the data collection and analysis to an educational level, only the data related to the

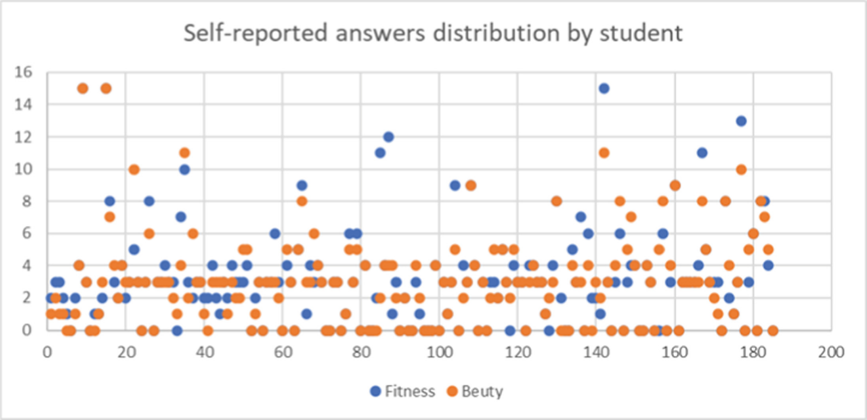


Fig. 3. The distribution of the score from the self-reported questions (Y axis) performed by each of the 184 students (X axis)

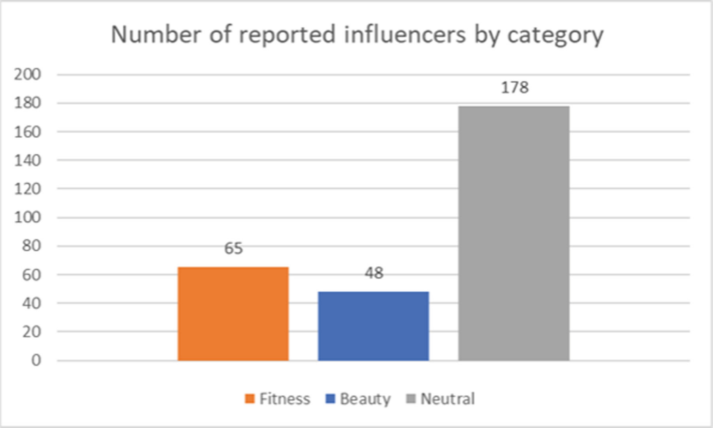


Fig. 4. Number of reported influencers by category

predefined profiles were considered. Students viewed photos 99 times, and opened the preloaded profiles 479 times, the predefined influencers were followed 91 times, 15 students left comments, and 43 liked a particular photo. The study of the xAPI data showed that 68.6% of students interacted at least one time with the fitness content related, 35.2% with the beauty, and 46% with the neutral. As shown in Fig. 5, the footprint that the students left behind was diverse which could be an indicator that students were interacting with the content following their personal preferences. In Fig. 6, the average of the interaction scores for each role shows that the most dominant category is the fitness aligned with the results of the first approach. Although the standard deviation of the answer is high, fitness is equal to 24,4 beauty 5,39 and neutral 12,02 this shows that not all the students had the same amount of influence by a stereotype.

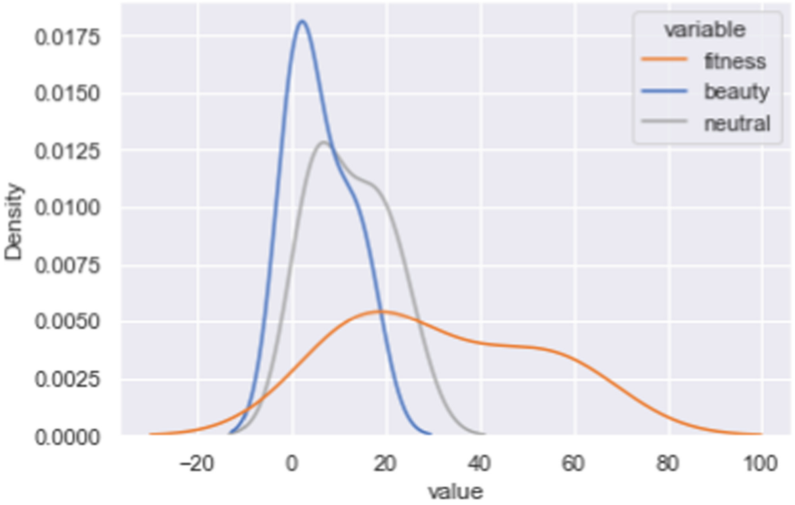


Fig. 5. The distribution of the score (KDE) for the three categories. Score value (X axis), density (Y axis)

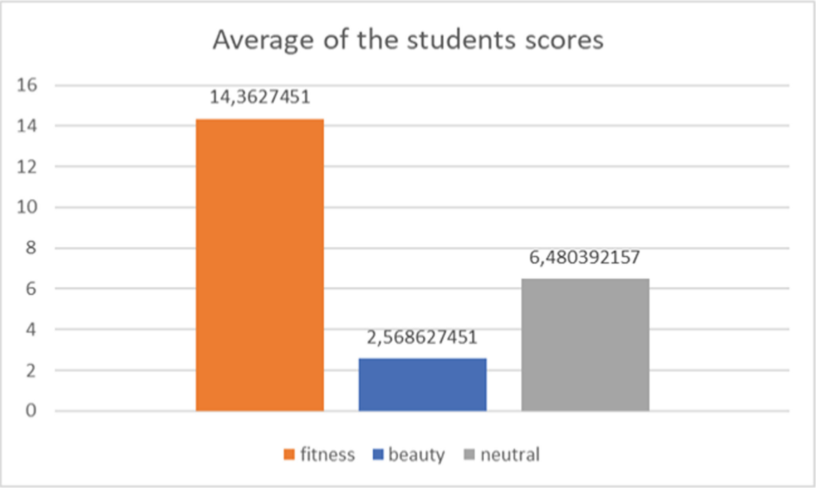


Fig. 6. Average score of the 102 students in the three categories

4 Discussion

The challenge of identifying students' body image stereotypes in an educational platform lies in understanding the user's interests. The two first approaches collected self-reported data from the students at the start of the first session. The first approach asked students to report their preferred source of information for the different types of content. In total, 142 students selected at least one source providing insights into their social media activities. However, this approach had its limitations as not all students selected an answer and there was not an option for neutral profiles. In addition, a few students did not pay much attention to the questions and selected the same answer for all the sources of information.

The second approach prompted the students to indicate the top three Instagram profiles they follow daily. This approach had a different set of limitations primarily on data analysis. Students reported personal profiles that did not reach the threshold of 10000 followers, the profile names were not written correctly, and finally, there was a need to manually identify the profile types during the data analysis, making a future automated analysis difficult. However, the list of influencers can be used as an input for the creation of more realistic predefined profiles in the educational social network.

The third approach of xAPI gathered the most data points as students performed 723 interactions with the predefined content in a varied way. Students interacted with the predefined content naturally as they were not instructed to interact with it, a potential indicator of their real interest. However, only 102 students out of the 184 interacted with the predefined content.

In all three approaches, the narrative of fitness was the most dominant one showing a form of consistency across the different approaches. With each approach having its own sets of strengths and limitations (see Table 1) its selection merely depends on the type of study and system that will be implemented.

Table 1. Strengths and limitations of the three approaches

Approach	Strengths	Limitations
Self-reported questions	- Number of answers, as questions can be marked as mandatory in a questionnaire	- Is not possible to know the validity of the answers and students' levels of attention - An integration between the survey software and the education platform is needed
Top influencers list	- Can be used to know the general interests of the group - Identify direct dangers hidden behind the profiles that students interact with	- Difficult to integrate, the classification of the influencers has to be done manually
xAPI	- Unsupervised and more natural interaction with the content - Easily integrated in the educational platform	- Is not guaranteed that the students will interact with the predefined content

However, computer-based detection techniques can be crucial for new learners because information is initially insufficient to build appropriate learner profiles [32]. Some computer-based detection techniques require large amounts of training data to achieve accurate trait identification [33]. Therefore, some researchers use a hybrid technique, which combines two or more techniques (either a mix of both questionnaire and computer-based techniques or a combination of computer-based techniques) [34].

These mix of questionnaires and computer-based trait estimation can be used to adapt the collaborative learning task structure (changing the order of the activities based on the student) or to modify the social structure of future activities (adjusting the group formation).

5 Conclusion

A comparison between two self-reported approaches and the use of xAPI has revealed a set of strengths and limitations for each approach. With the preliminary results showing a similar tendency of dominant stereotype preference amongst the students, the selection of the most appropriate approach depends on the type of study that will be conducted.

The use of xAPI has shown to be an adequate approach for an automated system to track students' behavior. The students successfully interacted with predefined influencers on a social media educational platform and their interactions matched their self-reported top followed influencer types. Therefore, in a future study, the implementation of adaptive counter-narratives can be achieved by tracking the student's digital footprint using xAPI and a counter-narrative allocation algorithm. The self-reported approach using indirect questions about the studied categories also showed good results and can be used as a complementary way to detect the dominant stereotype for the students that do not interact with the predefined profiles and do not generate xAPI data.

In future studies, we plan to use the xAPI approach to provide an adaptive learning experience with the use of the counter-narratives, by creating an intervention using the Jigsaw CLFP and grouping students based on their dominant category. Also, we plan to extend this work by implementing more types of counter-narratives based on body image perception, body image preoccupation, and dissatisfaction.

Acknowledgements. This work has been partially supported by grants PID2020-112584RB-C33, MDM-2015-0502 funded by MICIN/AEI/https://doi.org/10.13039/501100011033 and the Volkswagen Stiftung (Courage, ref. 95 566). D. Hernández-Leo (Serra Húnter) acknowledges the support by ICREA under the ICREA Academia programme.

References

1. de Lenne, O., Vandenbosch, L., Eggermont, S., Karsay, K., Trekels, J.: Picture-perfect lives on social media: a cross-national study on the role of media ideals in adolescent well-being. Media Psychol. **23**(1), 52–78 (2020)
2. Fardouly, J., Vartanian, L.R.: Social media and body image concerns: current research and future directions. Curr. Opin. Psychol. **9**, 1–5 (2016). https://doi.org/10.1016/j.copsyc.2015.09.005
3. Ahadzadeh, A.S., Pahlevan Sharif, S., Ong, F.S.: Self-schema and self-discrepancy mediate the influence of Instagram usage on body image satisfaction among youth. Comput. Hum. Behav. **68**, 8–16 (2017). https://doi.org/10.1016/J.CHB.2016.11.011
4. Marengo, D., Longobardi, C., Fabris, M.A., Settanni, M.: Highly-visual social media and internalizing symptoms in adolescence: the mediating role of body image concerns (2018).https://doi.org/10.1016/j.chb.2018.01.003

5. Verrastro, V., Liga, F., et al.: Fear the Instagram: beauty stereotypes, body image and Instagram use in a sample of male and female adolescents. Qwerty Open Interdiscip. J. Technol. Cult. Educ. **15**, 31–49 (2020). https://doi.org/10.30557/QW000021
6. Cash, T.F., Smolak, L. (eds.): Body Image: A Handbook of Science, Practice, and Prevention. Guilford Press (2011)
7. Saiphoo, A.N., Vahedi, Z.: A meta-analytic review of the relationship between social media use and body image disturbance. Comput. Hum. Behav. **101**, 259–275 (2019). https://doi.org/10.1016/j.chb.2019.07.028
8. Hou, Y., Xiong, D., Jiang, T., et al.: Social media addiction: its impact, mediation, and intervention. Cyberpsychol. J. Psychosoc. Res. Cyberspace (2019). https://doi.org/10.5817/cp2019-1-4
9. McLean, S., Wertheim, E., Masters, J., Paxton, S.: A pilot evaluation of a social media literacy intervention to reduce risk factors for eating disorders. Int. J. Eat. Disord. **50**, 847–851 (2017). https://doi.org/10.1002/eat.22708
10. Sánchez-Reina, J.R., Fuentes, C.B.: Comunicación De La Salud En La Campaña «Chécate, Mídete, Muévete». Representaciones y eficacia. Razón y Palabra **20**(94), 645–662 (2016)
11. Hernández-Leo, D., Theophilou, E., Lobo, R., Sánchez-Reina, R., Ognibene, D.: Narrative scripts embedded in social media towards empowering digital and self-protection skills. In: De Laet, T., Klemke, R., Alario-Hoyos, C., Hilliger, I., Ortega-Arranz, A. (eds.) EC-TEL 2021. LNCS, vol. 12884, pp. 394–398. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86436-1_42
12. New Media Consortium: NMC Horizon Report: 2018 Education Edition. Retrieved June (2018)
13. Hattie, J.: Visible Learning: A Synthesis of over 800 Meta-analyses Relating to Achievement. Routledge, London (2008)
14. Taylor, D.L., Yeung, M., Bashet, A.Z.: Personalized and adaptive learning. In: Ryoo, J., Winkelmann, K. (eds.) Innovative Learning Environments in STEM Higher Education. SpringerBriefs in Statistics, pp. 17–34. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-58948-6_2
15. Woolf, B.P.: Student modeling. Stud. Comput. Intell. **308**, 267–279 (2010). https://doi.org/10.1007/978-3-642-14363-2_13
16. Baiti, N.: Identification of personal traits in adaptive learning environment: systematic literature review. Comput. Educ. **130**, 168–190 (2019). https://doi.org/10.1016/j.compedu.2018.11.005. ISSN 0360-1315
17. Dillenbourg, P.: Split where interaction should happen-a model for designing CSCL scripts. In: Instructional Design for Effective and Enjoyable Computer-Supported Learning, pp. i–ii (2004)
18. Jermann, P., Dillenbourg, P.: Elaborating new arguments through a CSCL script. In: Andriessen, J., Baker, M., Suthers, D. (eds.) Arguing to Learn, pp. 205–226. Springer, Dordrecht (2003). https://doi.org/10.1007/978-94-017-0781-7_8
19. Amarasinghe, I., Hernández-Leo, D., Jonsson, A.: Data-informed design parameters for adaptive collaborative scripting in across-spaces learning situations. User Model. User-Adap. Inter. **29**(4), 869–892 (2019). https://doi.org/10.1007/s11257-019-09233-8
20. Fasihuddin, H., Skinner, G., Athauda, R.: Towards an adaptive model to personalise open learning environments using learning styles. In: Proceedings of International Conference on Information, Communication Technology and System (ICTS), pp. 183–188 (2014). https://doi.org/10.1109/ICTS.2014.7010580
21. Aslan, S., et al.: Students' emotional self-labels for personalized models. In: Proceedings of the Seventh International Learning Analytics & Knowledge Conference (LAK 2017), pp. 550–551. Association for Computing Machinery, New York (2017). https://doi.org/10.1145/3027385.3029452

22. Hidayat, A., Utomo, V.G.: Automatic detection of learning style in adaptive online module system. In: 2016 International Conference on Informatics and Computing (ICIC), pp. 94–98 (2016).https://doi.org/10.1109/IAC.2016.7905696
23. Verrastro, V., Fontanesi, L., Liga, F., Cuzzocrea, F., Gugliandolo, M.C.: Fear the Instagram: beauty stereotypes, body image and Instagram use in a sample of male and female adolescents. Qwerty **15**(1), 31–49 (2020). https://doi.org/10.30557/QW000021
24. Niemann, Y.F., Jennings, L., Rozelle, R.M., Baxter, J.C., Sullivan, E.: Use of free responses and cluster analysis to determine stereotypes of eight groups. Pers. Soc. Psychol. Bull. **20**(4), 379–390 (1994). https://doi.org/10.1177/0146167294204005
25. Butkowski, C.P., Dixon, T.L., Weeks, K.R., Smith, M.A.: Quantifying the feminine self(ie): gender display and social media feedback in young women's Instagram selfies. New Media Soc. **22**(5), 817–837 (2020). https://doi.org/10.1177/1461444819871669
26. Kitto, K., Cross, S., Waters, Z., Lupton, M.: Learning analytics beyond the LMS, pp. 11–15 (2015). https://doi.org/10.1145/2723576.2723627
27. Cooper, A.: Learning analytics interoperability-the big picture in brief. Learn. Anal. Community Exchange 1–7 (2014)
28. De Croon, R., Wildemeersch, D., Wille, J., Verbert, K., Vanden Abeele, V.: Gamification and serious games in a healthcare informatics context. In: Proceedings of 2018 IEEE International Conference on Healthcare Informatics, ICHI 2018, August, pp. 53–63 (2018). https://doi.org/10.1109/ICHI.2018.00014
29. Wang, Y., Wang, M.: Data acquisition model for online learning activity in distance English teaching based on xAPI. Int. J. Continuing Eng. Educ. Life Long Learn. **31**(1), 1–16 (2021)
30. Manso-Vazquez, M., Caeiro-Rodriguez, M., Llamas-Nistal, M.: An xAPI application profile to monitor self-regulated learning strategies. IEEE Access **6**, 42467–42481 (2018). https://doi.org/10.1109/ACCESS.2018.2860519
31. Davies, G., Ouellet, M., Bouchard, M.: Toward a framework understanding of online programs for countering violent extremism. J. Deradicalization **6**, 51–86 (2016)
32. Baldiris, S., Graf, S., Fabregat, R.: Dynamic user modeling and adaptation based on learning styles for supporting semi-automatic generation of IMS learning design. In: IEEE International Conference on Advanced Learning Technologies, pp. 218–220. IEEE Computer Society, July 2011
33. Lukasenko, R., Grundspenkis, J.: Adaptation of intelligent knowledge assessment system based on learner's model. In: Proceeding on the 16th International Conference on Information and Software Technologies, Kaunas, Lithuania (2010)
34. Normadhi, N.B.A., Shuib, L., Nasir, H.N.M., Bimba, A., Idris, N., Balakrishnan, V.: Identification of personal traits in adaptive learning environment: systematic literature review. Comput. Educ. **130**, 168–190 (2019)

Students' Basic Psychological Needs Satisfaction at the Interface Level of a Computer-Supported Collaborative Learning Tool

Eyad Hakami(✉), Khadija El Aadmi-Laamech, Lubna Hakami, Patricia Santos, Davinia Hernández-Leo, and Ishari Amarasinghe

Universitat Pompeu Fabra, Barcelona, Spain
{eyad.hakami01,lubna.hakami01}@estudiant.upf.edu,
{khadija.elaadmi,patricia.santos,davinia.hernandez-leo,
ishari.amarasinghe}@upf.edu

Abstract. Well-being has been considered an urgent vein of discussion in fields that intersect with Information and Communication Technologies. In this paper, we used a questionnaire adapted from the METUX (Motivation, Engagement, and Thriving in User Experience) model to explore how well a Computer-Supported Collaborative Learning (CSCL) tool's interface satisfy users' needs for competence, autonomy, and relatedness; and to test the instrument's validity in a CSCL context. METUX provides scales grounded in Self-Determination Theory (SDT) allowing researchers to foster insights into how technology designs support or undermine psychological needs, boosting user well-being. 53 bachelor students represented the tool's users based on convenience sampling. Our findings showed that users may not perceive the autonomy construct in the tools' interface, taking a neutral stance toward aspects of competence and relatedness as well. The results indicate the need for design interventions to improve the interface's ease of use, and the components that facilitate interaction and feelings of being connected. Regarding the instrument, more work is needed to validate the use of METUX interface in CSCL, especially for the autonomy subscale. Also, more scales from METUX (e.g., adoption and task spheres of experience) are needed to be included in the future for a fuller validation.

Keywords: Well-being · Computer-supported collaborative learning · Self-determination theory · METUX

1 Introduction

The satisfaction of three basic psychological needs—*competence* (the sense of being capable and effective), *autonomy* (feeling self-governed and self-endorsed) and *relatedness* (feeling connected and interacting)—has been shown to be critical to both motivation and well-being in the field of psychology [1]. According to the Self-Determination Theory (SDT) [2], the satisfaction of these three needs is a universal prerequisite for psychological well-being. SDT theorists [2–5] consider these needs as broad motivational

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 218–230, 2022.
https://doi.org/10.1007/978-3-031-20218-6_15

inclinations that function throughout life domains and argue that satisfaction of all three needs, as opposed to only one or two, is crucial for well-being [6]. In education, SDT posits that students' intrinsic motivation is rooted in having their basic psychological needs met [3]. Students are actively motivated to engage in learning tasks when pedagogical design appropriately satisfies these psychological needs [7]. The majority of SDT studies in this regard have investigated how the three needs are fulfilled in traditional face-to-face learning [8, 9], with some exceptions discussing SDT in online and digital learning contexts [7, 10]. One current direction of SDT research concerns the potential and challenges associated with the use of technologies in education [11]. More SDT research, according to [11], will undoubtedly be looking at not only how technology-enhanced learning can be designed to motivate engagement and learning [12], but also how teachers and students can be motivated to embrace technology as a tool for learning [13, 14]. In collaborative learning, sense of relatedness is particularly relevant due to the great amount of social interaction involved in collaborative settings. For example, a study by [15] showed that students' sense of relatedness to peers and teachers predicted their engagement level in collaborative writing using wikis.

The past decade has seen a rise in interest in human-centred design, where scholars and practitioners alike have struggled to translate the desire to design for human flourishing and well-being into clear and practical practice. The three basic needs can be utilised as inspirations or parameters to evaluate and enhance a design [13, 16]. Designing with users' psychological needs in mind (i.e., their desire to feel competent and autonomous, as well as their need to feel connected to others) is a key component of the SDT approach [13]. The notion of needs satisfaction implies that designers are required to understand users' expectations regarding the needs and adjust the design to meet those expectations [13]. For example, [17] applied SDT to understand what the three psychological needs entail in conversational agents' experiences. That study obtained insights into users' perceptions and expectations on the three needs, enabling the development of informative recommendations for fulfilling the needs in the design of conversational agents [17].

In this paper, we apply METUX TENS-Interface [13], a measure driven from SDT-based questionnaires, to explore students' perceptions on the extent to which their basic psychological needs are satisfied at the interface level of using PyramidApp, a computer-supported collaborative learning (CSCL) tool. PyramidApp is a web-based tool that enables teachers to design and implement CSCL scripts based on the Pyramid pattern [18]. Within the tool, students engage in collaboration following a Pyramid structure. Students are automatically allocated into small groups first and later into larger groups, facilitating them to reach a consensus to the given task at the end of the script. A teacher-facing dashboard is built into the tool to support teachers in orchestrating collaboration [19]. This work aims at exploring whether the three basic psychological needs are covered by the tool; and validating the used instrument in the tool's context for the purposes of continuous data collection and evaluation. We posit that the use of METUX TENS-Interface questionnaire in CSCL can provide meaningful insights about user autonomy, competence and relatedness; and therefore, inform the design processes in these regards.

The rest of this paper is organised as follows: We review the research context and the studied tool. Then we clarify the methods followed in this research, explaining the previous work and METUX model with a focus on the TENS-Interface questionnaire.

Then we test the scales' validity, visualise and discuss the findings and conclude the paper by describing the implications of design and the future direction of this work.

2 Research Context

2.1 Self-determination Theory (SDT)

Self-determination theory (SDT) posits that basic psychological needs for autonomy, competence, and relatedness must be satisfied for an individual, at all ages, to develop a sense of growth, integrity, and well-being [4, 20]. Experiencing the feeling of effectiveness and mastery is central to the concept of competence. As one effectively completes tasks and encounters opportunities to apply skills and knowledge, this need is fulfilled. Feelings of inefficiency and failure are common responses to competence frustration. Autonomy is the experience of voluntary action, and is satisfied when one's behaviours, thoughts, and feelings are self-endorsed and authentic. When frustrated, one feels pressure, conflict, and being pushed in an undesired direction. Relatedness is the experience of bonding and care, and it is satisfied by feeling connected to others. Relatedness frustration comes with a feeling of being socially isolated and excluded [see 1–20]. There is sufficient evidence from SDT [21–23] that a learning environment that satisfies students' need for autonomy, competence, and relatedness is essential for learners' self-determination and self-regulation. Students' intrinsic motivation, autonomous self-regulation, along with the quality of their performance, are influenced by the extent to which their basic psychological needs are satisfied in their learning environments [1, 4].

2.2 Pyramid Pattern Based CSCL Activities

Computer-Supported Collaborative Learning (CSCL) is an interdisciplinary field of research that aims to investigate how learners engage in collaboration with the help of computers [24]. Although CSCL provides opportunities to connect peers with the use of computers, there is no guarantee that every CSCL situation may create opportunities for productive interactions and therefore learning. To this end, scripts had been proposed as a way to structure collaboration by providing guidance and instructions to students on how to interact during collaboration in Technology Enhanced Learning scenarios [25, 26]. These 'scripts' are known as Collaborative Flow Patterns (CLFPs). Some of the well-known examples of CLFPs include Pyramid, Jigsaw, Think-Pair-Share (TPS), and Thinking Aloud Pair Problem Solving (TAPPS) [27].

Different CLFPs are shaped by the pedagogical rationale and constraints defined by CLFPs themselves [28]. For instance, Pyramid CLFP integrates activities occurring at multiple social levels. First learners will study a given problem individually to propose an initial solution. Learners then join in small groups, usually in pairs to discuss their solutions, and to propose a shared solution at the small group level. The discussion and negotiation will repeat in growing sizes of groups following a Pyramid structure until the whole group reaches a common solution to the given problem. Structuring collaboration according to this pattern provides several educational benefits to students. For instance, it provides equal opportunities for students to express their solutions, to negotiate with their peers, and also as the interactions accumulate across Pyramid levels it promotes positive interdependence. In this study, a tool called PyramidApp that implements a particularisation of the Pyramid pattern has been used to deploy CSCL activities. The tool provides an activity authoring space and a teacher-facing dashboard for the teachers and an activity enactment space for students. The teacher-facing dashboard not only provided a real-time overview of collaboration but also consisted of different controls, e.g., activity pause-resume, increasing time, and an alerting mechanism that informed critical moments of collaboration to the teachers to support their orchestration actions.

2.3 PyramidApp

PyramidApp is a web-based tool that facilitates the implementation of the Pyramid pattern-based collaborative learning activities [19, 28]. The tool is composed of three main components namely: a) activity authoring/design space; b) activity enactment space and c) activity regulation space. As shown in Fig. 1 first in the activity design stage teachers are required to configure several design elements related to the group activity such as the number of students in class, duration of the script phases, and group size. Once designed the activity can be published to generate an automatic URL that can later be shared with students for enactment. Students can use their mobile phones, tables or laptops to join the activity. The tool also provides a teacher-facing dashboard through which the teacher can monitor collaboration and intervene as required.

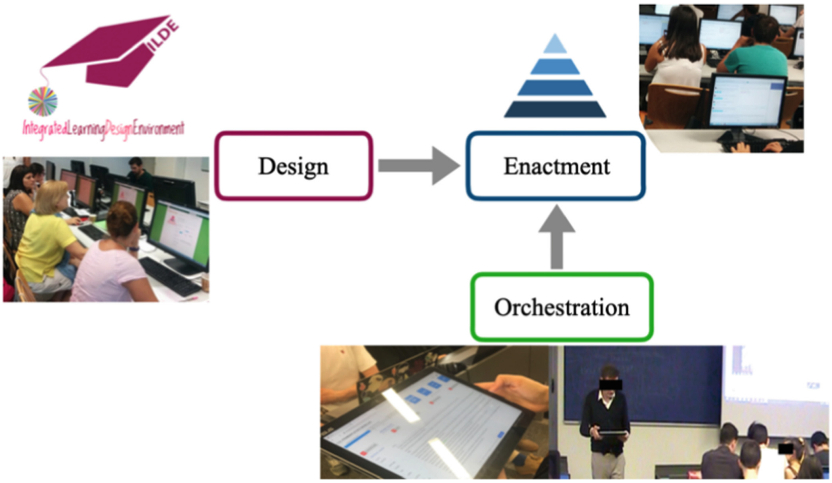
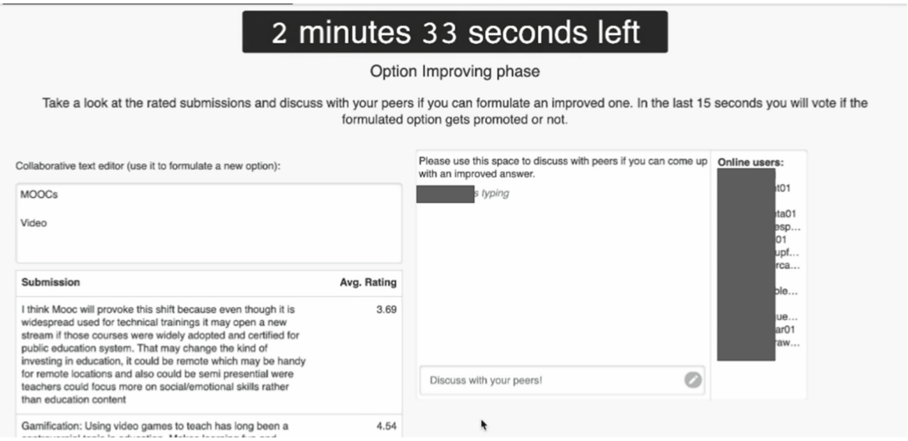


Fig. 1. Different components of PyramidApp

Within the PyramidApp collaboration is structured following a Pyramid structure. After login into the tool, students are required to enter an individual answer to the given problem. At the end of the individual answer submission stage students are randomly allocated into groups where they get an opportunity to see the answers submitted by the fellow group members. At the group levels, students are expected to evaluate the answers from peers. At the end of the voting phase students moved into an option improving phase (see Fig. 2). In this phase students had access to the integrated chat to engage in discussion with peers and a collaborative text editor (see top-left in Fig. 2) that provided a space for students to write an improved option or to reformulate existing options collaboratively. Students were also shown the average ratings received for each option at the previous rating level (see bottom-left in Fig. 2). At the end of the option improving stage students were promoted to agree on the newly formulated option or to promote the previous answers to further evaluate in the next larger group levels (Fig. 3). Also, all the groups are merged to formulate larger groups. Again, in the larger groups within an individual option evaluation stage students first evaluated the selected options from the previous small group levels individually, then engaged in the option improving stage as discussed earlier. At the end of the activity the selected answers are presented to the students.



2 minutes 33 seconds left

Option Improving phase

Take a look at the rated submissions and discuss with your peers if you can formulate an improved one. In the last 15 seconds you will vote if the formulated option gets promoted or not.

Collaborative text editor (use it to formulate a new option):

MOOCs

Video

Submission	Avg. Rating
I think Mooc will provoke this shift because even though it is widespread used for technical trainings it may open a new stream if those courses were widely adopted and certified for public education system. That may change the kind of investing in education, it could be remote which may be handy for remote locations and also could be semi presential were teachers could focus more on social/emotional skills rather than education content	3.69
Gamification: Using video games to teach has long been a	4.54

Please use this space to discuss with peers if you can come up with an improved answer.

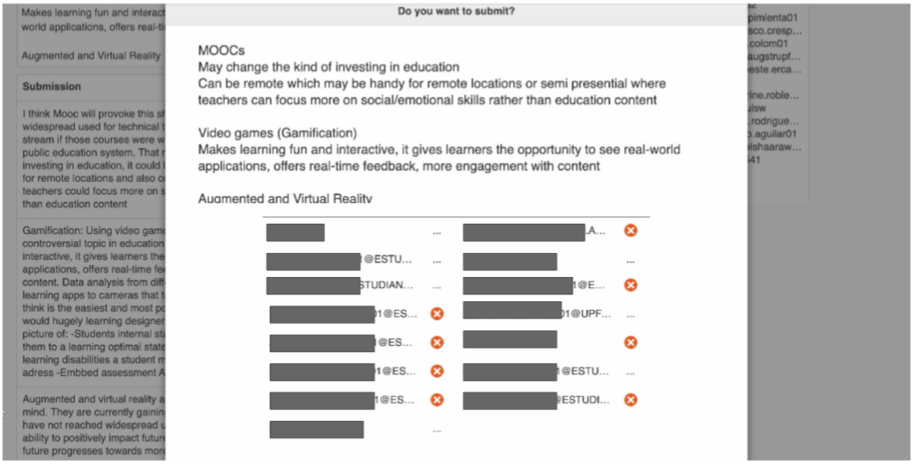
[redacted] typing

Discuss with your peers!

Online users:

[redacted]

Fig. 2. User interface of the PyramidApp, answer improving space (left), discussion space (right)



Makes learning fun and interact
world applications, offers real-ti

Augmented and Virtual Reality

Submission

I think Mooc will provoke this sh
widespread used for technical t
stream if those courses were w
public education system. That r
investing in education, it could b
for remote locations and also c
teachers could focus more on s
than education content

Gamification: Using video game
controversial topic in education
interactive, it gives learners the
applications, offers real-time fe
content. Data analysis from diff
learning apps to cameras that t
think is the easiest and most po
would hugely learning designer
picture of: -Students internal st
them to a learning optimal state
learning disabilities a student m
adress -Embbed assessment A

Augmented and virtual reality a
mind. They are currently gainin
have not reached widespread u
ability to positively impact futur
future progresses towards mor

Do you want to submit?

MOOCs
May change the kind of investing in education
Can be remote which may be handy for remote locations or semi presential where teachers can focus more on social/emotional skills rather than education content

Video games (Gamification)
Makes learning fun and interactive, it gives learners the opportunity to see real-world applications, offers real-time feedback, more engagement with content

Auamented and Virtual Realitv

[redacted] ... [redacted] .A... ⊗
[redacted] @ESTU... ... [redacted] ...
[redacted] STUDIAN... ... [redacted] 1@E... ⊗
[redacted] 1@ES... ⊗ [redacted] 01@UPF... ...
[redacted] @ES... ⊗ [redacted] ⊗
[redacted] 1@ES... ⊗ [redacted] 1@ESTU... ...
[redacted] 1@ES... ⊗ [redacted] @ESTUDI... ⊗
[redacted] ...

pimienta01
sco.cresp...
.colom01
augstrupf...
este.erca...
ine.roble...
lsw
.rodrigue...
.aguilar01
lshaaraw...
41

Fig. 3. Agreeing on newly formulated options

3 Methods

3.1 Previous Work

The inquiry in this paper belongs to a broader framework where an evaluation process guided by the IEEE P7010-2020 Well-being Impact Assessment (WIA) is applied to evaluate the well-being impact of PyramidApp on its users and stakeholders. As a methodology, WIA consists of five activities: 1) Internal, user, and stakeholder analysis, 2) well-being indicators dashboard creation, 3) data collection plan and data collection, 4) well-being data analysis and use of well-being indicators data, and 5) Iteration [29]. This paper is related to the third activity, aiming at collecting data that can be used to enhance the studied tool's digital well-being. Two of the tool's developers and a sample of the tool's users and stakeholders had participated in surveys and interviews to reflect on a wide range of well-being indicators distributed to multiple well-being domains. The findings discussed possible impacts on the well-being of students and teachers in the areas of life satisfaction, affect (stress), psychological state (sense of capability), community (sense of belonging), education (learning), human settlement (ICT skills), and work (support from peers) [30].

3.2 METUX TENS-Interface

METUX (Motivation, Engagement, & Thriving in User Experience) is a model for bridging Self Determination Theory (SDT) to technology design practice [13]. METUX can be used to evaluate technologies with respect to well-being impact when well-being in this context refers to the "optimal psychological functioning and experience" [31]. The METUX model centres on the well-researched claim [1] that human psychological well-being is mediated by three key constructs: Autonomy (feeling agency, acting in accordance with one's goals and values), Competence (feeling able and effective); and Relatedness (feeling connected to others, a sense of belonging) [13].

METUX proposes that in order to address well-being, psychological needs must be considered within five different spheres of analysis including: at the point of technology *adoption*, during interaction with the *interface*, as a result of engagement with technology-specific *tasks*, as part of the technology-supported *behavior*, and as part of an individual's *life* overall [13]. The data we collect and analyze in this paper is limited to the interface sphere by applying the TENS-Interface questionnaire to a sample of a CSCL tool's student users. When students interact with a learning tool, the satisfaction of the basic psychological needs, via the user interface, predict usability, engagement with technology, and user satisfaction. On the other hand, poor interface usability will cause need-frustration which impacts both engagement and user well-being [13].

3.3 Procedures

A sample of the studied tool's users, 53 first year bachelor students who were enrolled to the same course at a Spanish university, was selected based on convenience sampling. The participants were asked to rate their level of agreement to 15 items using a 5-point Likert scale (1 = Do Not Agree, 5 = Strongly Agree). Each key construct (e.g., competence) was measured through five items. All items are weighted equally in scoring, and reverse-scored items are reverse scored. The participants filled the questionnaire after they finished a task facilitated by the tool. All the participants had used the tool to complete collaborative learning tasks at least on three occasions by the time of filling the survey.

4 Findings

4.1 Validity Statistics

The measures introduced in METUX were externally validated by the model's developers [13], who carried out a pilot validation study in which 400 participants (100 for each of four technologies) were asked to fill out each METUX questionnaire in reference to one of four possible technologies: Facebook, Google Docs, a music streaming service and a fitness band. Results showed satisfactory to good internal consistency for all questionnaires with alphas for subscales ranging from 0.66 to 0.88. Furthermore, some initial support for the METUX model in higher education was provided by [32], who urged the need for additional validation work to improve the scale that measures need-satisfaction in the interface and task spheres of experience.

We conducted a validity analysis on the TENS-Interface questionnaire comprising five items for each subscale to test their validity in a CSCL context. Cronbach's alpha showed that the competence and relatedness subscales reached good internal consistency levels, $\alpha = 0.85$ and $\alpha = 0.80$ respectively. The autonomy subscale failed to reach the minimum accepted value of Cronbach's alpha, which was found at $\alpha = 0.67$ [13] and had a questionable internal consistency of $\alpha = 0.63$.

Inter-item correlations and item-total correlations were calculated for the autonomy subscale to identify problematic items. Most items appeared to be problematic in this subscale, resulting in low inter-correlations and slight decrease in the Cronbach's alpha if the item was deleted. The one exception to this was the third item (i.e., I feel pressured by the tool), which would significantly decrease the Cronbach's alpha if it was deleted and had a higher item-total correlation and more consistently higher inter-item correlations (Tables 1 and 2).

This outcome aligns with the results from the initial analysis conducted by the tool's creators to evaluate its overall well-being impact [30]. The tool had been found impactful on psychological well-being in the sense of capability, social well-being in the sense of belonging, and affective well-being in the sense of stress. The indicator of autonomy had not been found relevant in earlier stages of this evaluation process [30].

4.2 Scale Statistics

The responses of each participant to each 5-item scale were combined by calculating the average score of each participant, then the average score of each scale. The analysis of the participants' responses to the TENS-Interface questionnaire showed that competence was the most satisfied need in the interface of the studied tool (Mean = 3.63), followed by autonomy (Mean = 3.15) and relatedness (Mean = 2.96) (Table 3).

Table 1. Inter-item correlations of autonomy subscale

	The tool provides me with useful options and choices	I can get the tool to do the things I want it to	I feel pressured by the tool	The tool feels intrusive	The tool feels controlling
The tool provides me with useful options and choices	1	0.63	0.30	−0.007	0.008
I can get the tool to do the things I want it to	0.63	1	0.21	0.04	−0.03
I feel pressured by the tool	0.30	0.21	1	0.50	0.50
The tool feels intrusive	−0.007	0.04	0.50	1	0.38
The tool feels controlling	0.008	−0.03	0.50	0.38	1

Table 2. Item-total correlations of autonomy subscale

Item	Item-total correlation	Cronbach alpha if item deleted
The tool provides me with useful options and choices	0.59	0.61
I can get the tool to do the things I want it to	0.55	0.62
I feel pressured by the tool	0.81	0.44
The tool feels intrusive	0.60	0.59
The tool feels controlling	0.61	0.61

Table 3. Descriptive statistics of each subscale

Scale	No. of items	α	n	Mean	Std
Competence	5	0.85	53	3.63	0.79
Autonomy	5	0.63	53	3.15	0.64
Relatedness	5	0.80	53	2.96	0.73

4.3 Visualization

In order to have a global overview of the data, we visualised it in a compact representation through different colours in a percentile system, making it easier to visually digest and compare (Figs. 4, 5 and 6).

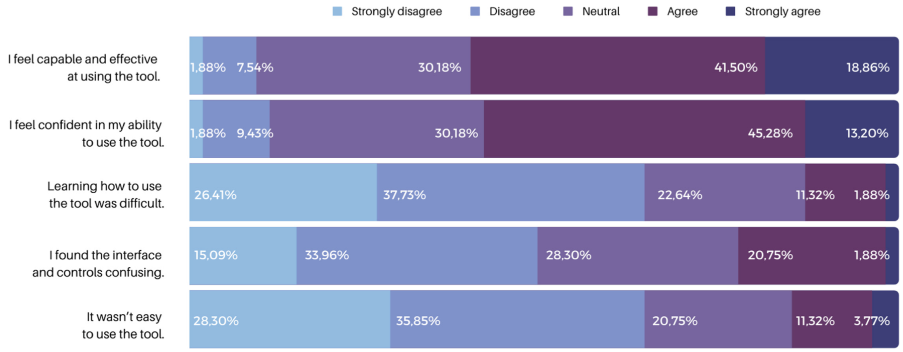


Fig. 4. Competence

5 Discussion and Future Work

SDT research and applications have grown significantly over the past two decades, with diverse interests in the relationship between the theory and practice in educational

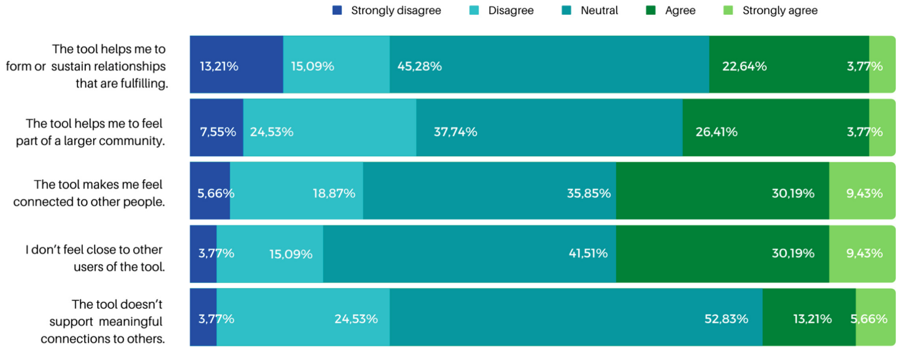


Fig. 5. Relatedness

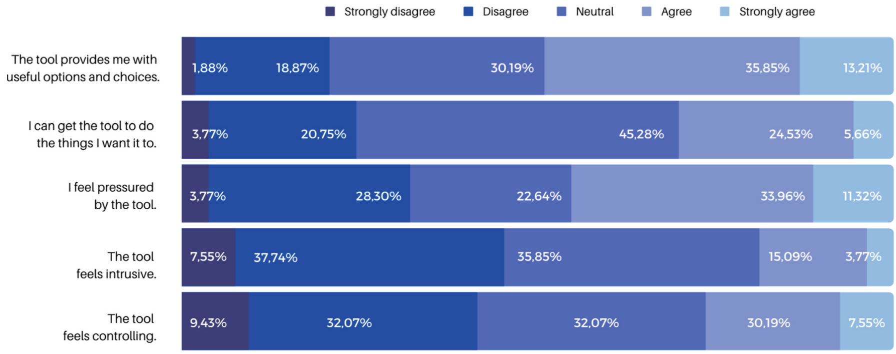


Fig. 6. Autonomy

contexts. In this paper, we explore how students' basic psychological needs for autonomy, competence, and relatedness are supported by the interface of a CSCL tool. The responses of 53 students who used the tool to complete collaborative learning tasks reveal that the value of autonomy is not as well defined as competence and relatedness in the interface of the studied tool. The internal consistency of the autonomy scale was questionable ($\alpha = 0.63$), indicating that the user may not clearly perceive this construct while dealing with the tool's interface. Some aspects of the relatedness construct (i.e., sustainable relationships and meaningful connections to others) are not well perceived as about half of the participants hold a neutral position towards them being supported in the tool's interface (Fig. 2). In addition, a third of the participants are neutral towards all of the competence aspects, indicating the need for design interventions to improve the interface's ease of use.

As for the TENS-Interface instrument itself, the low level of consistency in results we obtained in the Autonomy component might be due to the way the 5 questions are posed, since the questions can be perceived as generic, especially when the tool has a number of functionalities that we think should be evaluated separately for fuller insight on the true impact the interface has on the autonomy need. Thus, as part of our future work we propose to adapt the questions to each interface element or functionality, rather than compacting them all under the interface as a whole. As a step in this direction (specific to our tool), we propose to iterate the autonomy component of the TENS-Interface instrument, adapting it to the specific elements of the interface before proposing any tool design decisions in regard to autonomy need satisfaction.

On the other hand, related to the two remaining basic psychological needs, and based on the obtained results, since we find that competence was the most satisfied need in the interface of the studied tool (Mean = 3.63), we shift our focus to relatedness (Mean = 2.96), which was the least satisfied need. The design implications regarding the relatedness need are to be focused on tool components that facilitate students' interaction and feelings of being connected (i.e., chat, co-editing space and other collaborative components of the interface). The design improvements are to be evaluated by the same TENS-Interface questionnaire.

Overall, we presume that the TENS-Interface instrument requires further improvements before it can be utilized and applied to specific CSCL scenarios. We propose a first improvement in that regard: define the different functionalities of the interface first, then adapt the questions of the three components (autonomy, competence, relatedness) to each one of these functionalities, instead of relying solely on the interface as a whole. This will undoubtedly result in a longer questionnaire, but the results will be just as specific and detailed. Another positive aspect is that there will be more clarity on which components of the interface truly fulfil the three needs and which ones do not.

Acknowledgments. This work has been partially supported by grant PID2020-112584RBC33 funded by MICIN/AEI/10.13039/501100011033. D. Hernández-Leo (Serra Húnter) acknowledges the support by ICREA under the ICREA Academia program. P. Santos acknowledges the support by the Spanish Ministry of Science and Innovation under the Ramon y Cajal programme. E. Hakami acknowledges the grant by Jazan University, Saudi Arabia.

References

1. Ryan, R.M., Deci, E.L.: Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. Guilford Press, New York (2017)
2. Deci, E.L., Ryan, R.M.: Intrinsic Motivation and Self-Determination in Human Behavior. Plenum Press. (1985). (Taylor & Francis Online)
3. Ryan, R.M.: Psychological needs and the facilitation of integrative processes. J. Pers. **63**, 397–427 (1995)
4. Ryan, R.M., Deci, E.L.: Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp. Educ. Psychol. **25**, 54–67 (2000)
5. Ryan, R.M., Deci, E.L.: Self-determination theory and the role of basic psychological needs in personality and the organization of behavior. In: John, O.P., Robins, R.W., Pervin, L.A. (eds.) Handbook of Personality: Theory and Research, 3rd edn, pp. 654–678. Guilford, New York (2008)
6. Church, A.T., et al.: Need satisfaction and well-being: testing self-determination theory in eight cultures. J. Cross Cult. Psychol. **44**(4), 507–534 (2013)
7. Hsu, H.-C., Wang, C.V., Levesque-Bristol, C.: Reexamining the impact of self-determination theory on learning outcomes in the online learning environment. Educ. Inf. Technol. **24**(3), 2159–2174 (2019). https://doi.org/10.1007/s10639-019-09863-w

8. Lietaert, S., Roorda, D., Laevers, F., Verschueren, K., De Fraine, B.: The gender gap in student engagement: the role of teachers' autonomy support, structure, and involvement. Br. J. Educ. Psychol. **85**(4), 498–518 (2015). https://doi.org/10.1111/bjep.12095
9. Roorda, D.L., Koomen, H.M., Spilt, J.L., Oort, F.J.: The influence of affective teacher–student relationships on students' school engagement and achievement: a meta-analytic approach. Rev. Educ. Res. **81**(4), 493–529 (2011). https://doi.org/10.3102/0034654311421793
10. Chiu, T.K.F.: Applying the self-determination theory (SDT) to explain student engagement in online learning during the COVID-19 pandemic. J. Res. Technol. Educ. **54**(sup1), S14–S30 (2022). https://doi.org/10.1080/15391523.2021.1891998
11. Ryan, R.M., Deci, E.L.: Intrinsic and extrinsic motivation from a self-determination theory perspective. Definitions, theory, practices, and future directions. Contemp. Educ. Psychol. **61**, 101860 (2020). https://doi.org/10.1016/j.cedpsych.2020.101860
12. Ryan, R.M., Rigby, C.S.: Motivational foundations of game-based learning. In: Plass, J.L., Mayer, R.E., Homer, B.D. (eds.) Handbook of Game-Based Learning, pp. 153–176. The MIT Press, Cambridge (2019)
13. Peters, D., Calvo, R.A., Ryan, R.M.: Designing for motivation, engagement and wellbeing in digital experience. Front Psychol **9**, 797 (2018)
14. Sørebø, Ø., Halvari, H., Gulli, V.F., Kristiansen, R.: The role of self-determination theory in explaining teachers' motivation to continue to use e-learning technology. Comput. Educ. **53**, 1177–1187 (2009)
15. Law, W., King, R., Notari, M., Cheng, E., Chu, S.: Why do some students become more engaged in collaborative wiki writing? The role of sense of relatedness. In: Proceedings of the International Symposium on Open Collaboration (OpenSym 2014), pp. 1–6. ACM, New York (2014). https://doi.org/10.1145/2641580.2641603
16. Calvo, R., Peters, D.: Positive Computing: Technology for Wellbeing and Human Potential. MIT Press, Cambridge (2014)
17. Yang, X., Aurisicchio, M.: Designing conversational agents: a self-determination theory approach. In: CHI Conference on Human Factors in Computing Systems (CHI 2021), Yokohama, Japan, 08–13 May 2021, 16 p. ACM, New York (2021). https://doi.org/10.1145/3411764.3445445
18. Hernández-Leo, D., Villasclaras-Fernandez, E.D., Asensio-Perez, J.I., Dimitriadis, Y.A., Symeon, R.: CSCL scripting patterns: hierarchical relationships and applicability. In: Proceedings of the 6th IEEE International Conference on Advanced Learning Technologies (ICALT 2006), pp. 388–392 (2006). https://doi.org/10.1109/ICALT.2006.1652452
19. Manathunga, K., Hernández-Leo, D.: PyramidApp: scalable method enabling collaboration in the classroom. In: Verbert, K., Sharples, M., Klobučar, T. (eds.) EC-TEL 2016. LNCS, vol. 9891, pp. 422–427. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45153-4_37
20. Van den Broeck, A., Ferris, D.L., Chang, C.H., Rosen, C.C.: A review of self-determination theory's basic psychological needs at work. J. Manag. **42**(5), 1195–1229 (2016)
21. Grolnick, W.S., Ryan, R.M.: Parent styles associated with children's self-regulation and competence in school. J. Educ. Psychol. **81**, 43–54 (1989)
22. Deci, E.L., Vallerand, R.J., Pelletier, L.G., Ryan, R.M.: Motivation and education: the self-determination perspective. Educ. Psychol. **26**, 325–346 (1991)
23. Connell, J.P., Wellborn, J.G.: Competence, autonomy, and relatedness: a motivational analysis of self-system processes. In: Gunnar, M.R., Sroufe, L.A. (eds.) Self Processes and Development. The Minnesota Symposia on Child Psychology, vol. 23, pp. 43–77. Lawrence Erlbaum, Hillsdale (1991)
24. Stahl, G., Koschmann, T.D., Suthers, D.D.: Computer supported collaborative learning: a historical perspective. In: Cambridge Hand-Book of the Learning Sciences, pp. 409– 426. (2006)

25. Hernández-Leo, D., Asensio-Pérez, J.I., Dimitriadis, Y., Villasclaras Fernández, E.D.: Generating CSCL scripts: from a conceptual model of pattern languages to the design of real scripts. In: Technology-Enhanced Learning: Design Patterns and Pattern Languages, pp. 49–64 (2010)
26. Hernández-Leo, D., Asensio-Pérez, J.I., Dimitriadis, Y.: Computational representation of collaborative learning flow patterns using IMS learning design. Educ. Technol. Soc. **8**(4), 75–89 (2005)
27. Hernández-Leo, D., Asensio-Pérez, J.I., Dimitriadis, Y., Villasclaras, E.D.: Generating CSCL scripts: from a conceptual model of pattern languages to the design of real scripts. In: Goodyear P.; Retalis, S. (eds.) Technology-Enhanced Learning, Design Patterns and Pattern Languages, Series Technology-Enhanced Learning, pp. 49–64. Sense Publishers (2010)
28. Manathunga, K., Hernández-Leo, D.: Authoring and enactment of mobile pyramid-based collaborative learning activities. Br. J. Edu. Technol. **49**(2), 262–275 (2018). https://doi.org/10.1111/bjet.12588
29. IEEE: IEEE recommended practice for assessing the impact of autonomous and intelligent systems on human well-being. IEEE Std 7010-2020, pp. 1–96 (2020). https://doi.org/10.1109/IEEESTD.2020.9084219
30. Hakami, E., Hernández-Leo, D., Amarasinghe, I.: Understanding the well-being impact of a computer-supported collaborative learning tool: the case of PyramidApp. In: De Laet, T., Klemke, R., Alario-Hoyos, C., Hilliger, I., Ortega-Arranz, A. (eds.) EC-TEL 2021. LNCS, vol. 12884, pp. 373–378. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86436-1_38
31. Ryan, R.M., Deci, E.L.: On happiness and human potentials: a review of research on hedonic and eudaimonic well-being. Annu. Rev. Psychol. **52**(1), 141–166 (2001)
32. Jeno, L.M., Diseth, Å., Grytnes, J.A.: Testing the METUX model in higher education: interface and task need–satisfaction predict engagement, learning, and well-being. Front. Psychol. **12**, 631564 (2021). https://doi.org/10.3389/fpsyg.2021.631564

Facilitator Agent to Support Low-Resource Language Speakers in MT-Mediated Communication

Mondheera Pituxcoosuvarn(✉), Mizuki Motozawa, Yohei Murakami, and Shiyumi Yokote

Faculty of Information Science and Engineering, Ritsumeikan University, Shiga, Japan
mond-p@fc.ritsumei.ac.jp

Abstract. Machine translation (MT) has become useful in intercultural collaboration. However, for low-resource language (LRL) speakers, the translation accuracy possible might still be a burden to them. Previous studies showed that it is difficult for the minority and LRL speakers to participate in conversions. To solve this problem and create equal chance for team members to communicate, we aim at creating a facilitator agent that helps in supporting the LRL speakers or team members who might have problems joining the conversation. We achieve this by proposing the concept of a virtual facilitation agent that responds to and puts questions to the team members to support the discussion. Experiments on different facilitation strategies for discussion groups are conducted using our multilingual chat system.

Keywords: Facilitator agent · Machine translation · Intercultural collaboration

1 Introduction

In today's globalized society, the ability to understand and communicate with people and cultures from different countries is important. Machine translation (MT) technology can be used as a tool for communication across cultures allowing people from different countries to communicate with each other through MT. For example, in a summer school called "KISSY" organized by NPO Pangaea, children from various countries gathered and worked together using a multilingual communication chat system with embedded machine translation modules. By communicating with people from different countries, cultures, and languages, children can acquire the ability to understand and accept diverse values in a globalized society [9]. These kinds of collaboration are hindered by differences in culture and values, and there are unique ways of saying things in different countries. In order to understand these differences, it is important to strengthen communicate effectiveness. It is important to understand the other party's expressions

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 231–244, 2022.
https://doi.org/10.1007/978-3-031-20218-6_16

and thoughts, and also important to correctly transmit information to the other party.

Even if MT can help with the language barrier by providing translation, creating common ground between the parties still remains [12]. In addition, existing MT technology does not provide accurate translations for low-resource languages (LRL) which have fewer language resources, for example, having less bilingual data available to create MT services for those languages. As a result, LRL speakers are unable to actively participate in conversations including some participants of "KISSY" summer school. LRL speakers said fewer words than other language speakers [9].

This study aims to clarify effective communication strategies for facilitating LRL speakers in multilingual communication environments with different languages and cultural backgrounds.

The contribution of this paper is to define a facilitator agent whose behavior promotes LRL speech, and test its effectiveness through group discussions among people with different mother tongues.

2 Related Work

Researchers have been trying to develop and improve facilitator agents and conversation agents on different platforms [1,11]. For example, Ito et al. [6] used an automated facilitation agent to support crowd discussion on a discussion forum, while Kim et al. [7] developed a facilitation chatbot to be used in a chat application.

One of the existing support systems is the *listening dialogue system* [4]. This system offers chat dialog support with the goal being a listening dialogue system that can satisfy the user's desire for dialogue and maintain the cognitive function of elderly people. Other researchers have worked on the selection and generation of lexical responses that return idiomatic expressions in response to user utterances, responses that repeat parts of the utterances, and in-depth questions that inquire about the details of the content of the utterance [2].

Ishida et al. [4] published their work on the generation of self-disclosure responses, in which the system presents its own thoughts and information in response to the content of the user's utterance, in addition to in-depth questions, repetition responses, lexical responses, and evaluation responses, in order to create more natural and speech-friendly listening dialogues. In addition, they also proposed a method for judging whether each response is appropriate from a listener's point of view and selecting the appropriate type of response by using the results of speech recognition and focus analysis of the user's utterance and information such as captured responses as features.

Besides focusing on the listening agent, some researchers have focused on other features of the agent so replicate human agent performance as closely as possible. For example, Kitaoka et al. [8] studied the timing of responses to create a dialog system that can respond as reasonably as humans. In addition to the response time, replicating face gestures is also a factor. A group of researchers found that providing the agent with a face can enhance its interaction with humans in a conversation group [10].

The existing studies aimed at promoting conversation in monolingual communication, so the innovation of this research lies in its focus on supporting multilingual communication via MT.

3 Facilitator Utterance Design

3.1 Strategies

Based on the related research detailed in the previous section, we defined strategies that could be effective in supporting communication among LRL speakers.

The first strategy is to use utterances that request a summary of the discussion to facilitate LRL speakers' understanding of the content of the discussion and the meaning of others' utterances. The purpose of these utterances is to make it easier for LRL speakers to understand the situation of the discussion by asking high-resource language (HRL) speakers to briefly summarize the content of the discussion at that moment. The intention is to allow LRL speakers to understand the content of the discussion and the opinions of others, and to speak their opinions more easily.

Second, we define utterances that ask non-low-resource speakers to paraphrase utterance(s) in order to facilitate the low-resource speakers' understanding of the utterance(s). Utterances that are long might not be translated well by MT, so the facilitator agent will request non-low-resource speakers to paraphrase them briefly. This allows users to deepen their understanding of utterances that may be difficult for LRL speakers to understand. This approach is intended to simplify the message for LRL speakers.

Third, the facilitator agent sends utterances that respond to an LRL speaker's utterances. This strategy aims at making it easier for the LRL speaker to speak. This might help create an atmosphere in which LRL speakers find it easier to participate.

The fourth strategy is responding with utterances that return a positive response when an LRL speaker expresses an opinion. This type of utterance is an affirmative response, such as agreement, to an utterance by an LRL speaker, and its purpose is to create and encourage them to speak more actively.

3.2 Facilitator Agent Behavior

We designed the facilitator agent behavior based on the strategies defined in the previous section.

First, if the LRL speaker does not speak for a certain period of time, the facilitator requests a summary of the discussion. For example, the facilitator can tell everybody to "review the discussion so far" or "summarize the discussion". In our preliminary experiments, we found that it was effective to execute the command every three minutes, so we triggered command execution three minutes since the last utterance of the LRL speaker.

Next, the utterance requesting paraphrases from non-low-resource language (HRL) speakers is executed when an HRL speaker utters more than a certain number of characters in one utterance. Specifically, the facilitator agent can say, “Let’s summarize that in simple words” or “Please rephrase briefly”. In preliminary experiments, we found that if a user writes a message longer than 90 characters, there is a high probability that the content is difficult to understand. An utterance that responds to an utterance of an LRL speaker is executed when the LRL speaker speaks. Because it is a simple response, it is executed regardless of the content of the utterance. For example, the facilitator sends an utterance in the target LRL with messages such as “I see”, “uh-huh”, and “Is that so?”. The utterances that return a positive response when an LRL speaker expresses an opinion are executed when the LRL speaker expresses her or his thoughts and ideas. Specifically, they will not be executed in response to greetings, self-introductions, etc. The contents of the utterances include “I like it”, “It’s a nice idea,” and “I think it’s very good”. Because it is necessary to understand and judge the content of LRL speakers’ utterances, we used the Wizard Of Oz method, in which a human pretends to be a facilitator agent for this strategy, while the other strategies were executed using a virtual facilitator agent implemented as a chatbot.

A summary of all the strategies, purposes, execution conditions and execution methods is shown in Table 1.

Table 1. Facilitation Strategy

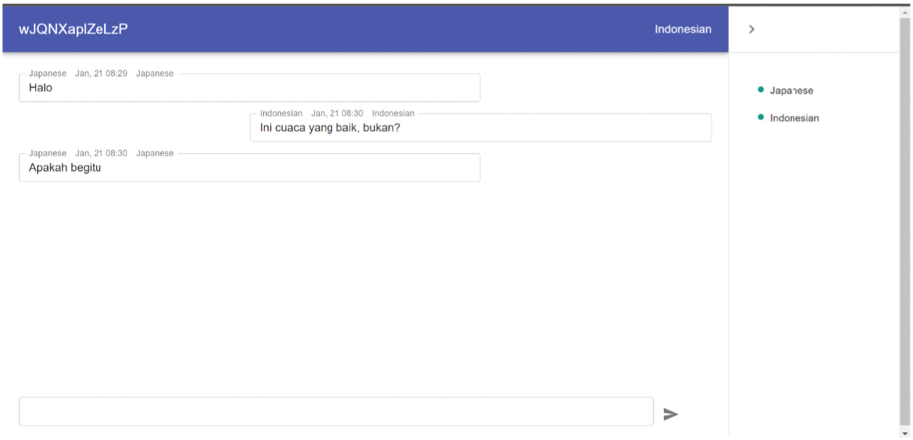
Strategy	Purpose (LRL speakers)	Condition	Method
Request for a summary	Promoting understanding	No utterance from LRL speakers for a certain period of time	Bot
Request to rephrase	Promoting understanding	Utterance from HRL speakers longer than a certain number of words	Bot
Responding to LRL utterance	Promoting utterances	Response to LRL speaker utterance	Bot
Positive response to LRL utterance	Promoting utterances	Response to LRL speaker opinion utterance	WOZ

4 Implementation

4.1 Overall System Configuration

LangridChat is a web application built on Django and React. Users can use this application to chat with other users in their preferred language. Currently, English, Japanese, Thai, Vietnamese, Indonesian, Nepali, Korean, Simplified Chinese, and Traditional Chinese are available. The server uses services from the Language Grid [5] to translate the input from the sender’s language to the languages selected by the receivers, which is then sent and displayed in the receivers’ language. The language can be changed by clicking on the current language at the top of the screen. The user can select a new language from the list.

The user can type a message in the text box at the bottom of the screen and then click on the arrowhead to send the message.



Japanese Jan, 21 08:29 Japanese
Halo

Indonesian Jan, 21 08:30 Indonesian
Ini cuaca yang baik, bukan?

Japanese Jan, 21 08:30 Japanese
Apakah begitu

Fig. 1. LangridChat interface

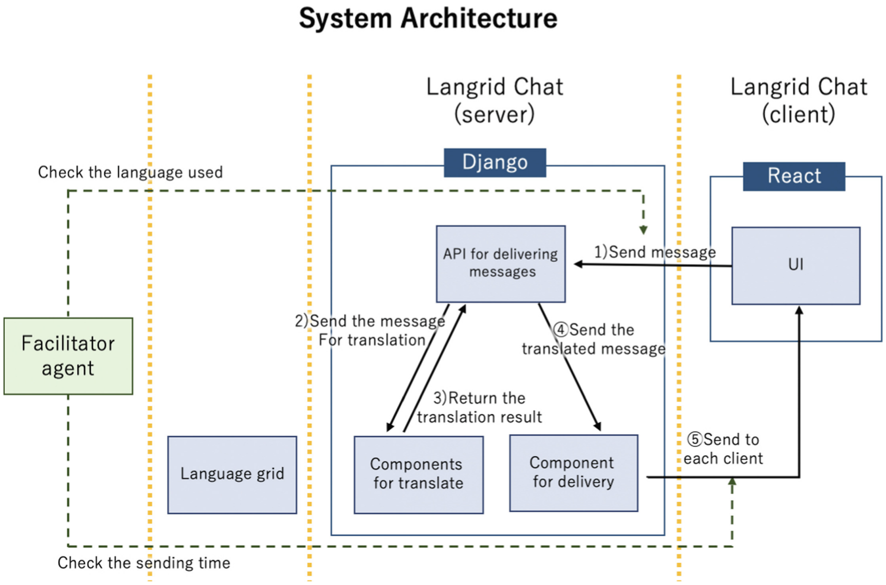


Fig. 2. System architecture

Figure 1 is a screenshot taken from a chat room with two users: a Japanese user and an Indonesian user. The message sent by the Japanese speaker in Japanese was translated into Indonesian and displayed on the screen of the Indonesian speaker.

The system is divided into two parts: one that runs on the server side and one that runs on the client-side, as shown in Fig. 2. The first is based on Django and includes an API for message delivery, a translation component, and a delivery component. Therefore, the server-side is responsible for creating chat rooms, translating and sending messages, recording chat logs, and retrieving user information. The latter has a React-based UI that records user information and displays sent messages. This front-end observes user behavior and responds on the client side.

4.2 Server-Side Implementation

The two types of utterances implemented on the server are paraphrase requests and responses to LRL speaker utterances.

Figure 3 displays the flowchart of sending a paraphrase request on the left and the flowchart of responding to LRL utterances on the right.

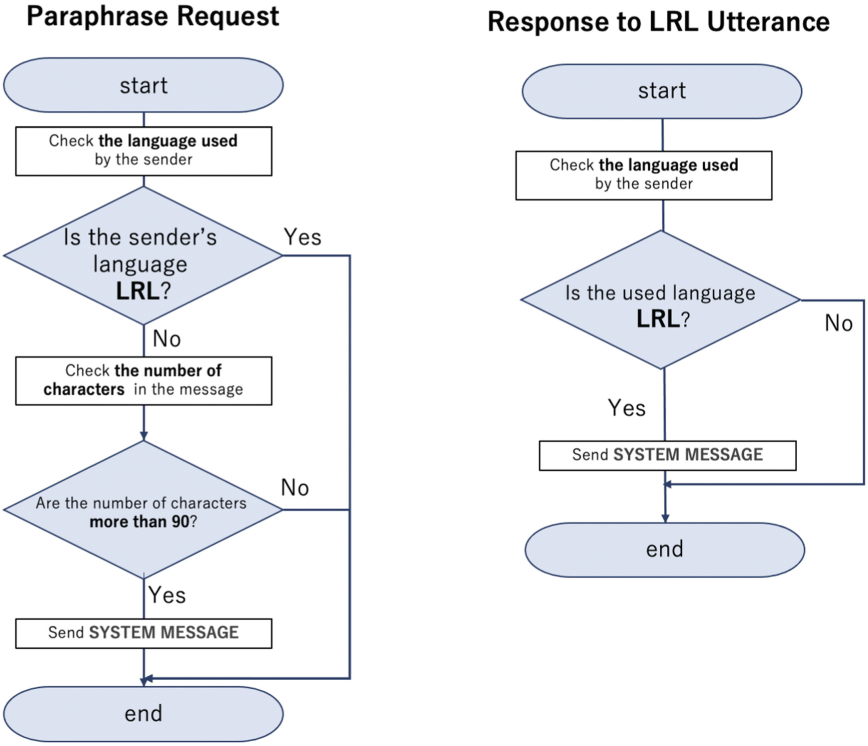

Paraphrase Request
start
Check the language used by the sender
Is the sender's language LRL?
Yes
No
Check the number of characters in the message
Are the number of characters more than 90?
No
Yes
Send SYSTEM MESSAGE
end
Response to LRL Utterance
start
Check the language used by the sender
Is the used language LRL?
No
Yes
Send SYSTEM MESSAGE
end

Fig. 3. Flowchart of paraphrase-request utterance generation (left) and responding to LRL utterances generation (right).

Since paraphrase requests are executed when a subject other than an LRL speaker sends 90 or more characters, it is necessary to observe the language used by the sender of the message and count the number of characters in the message. Since this information is exchanged on the server side, we implemented the system on the server side. If the language of the sender of the message is not Indonesian or Thai, the number of characters in the message is also checked. If the number of characters is more than 90, the server randomly sends a message with the user name of *SYSTEM MESSAGE*, saying “Let’s summarize in simple words” or “Please paraphrase briefly”.

An utterance that responds to an utterance of an LRL speaker is conditional on utterance type of the LRL speaker. In other words, the agent needs to observe the language used by the sender of the message, and as mentioned earlier, this information is controlled by the server side, so it needs to be implemented on the server. If the language of the sender of the message is Indonesian or Thai, the agent sends the message “seperti itu ya” (Is that so), “oh iya juga” (I see), or “Iya iya” (uh-huh).

4.3 Client-Side Implementation

Figure 4 shows the flowchart requesting a summary of the discussion. The utterance implemented on the client is an utterance requesting a summary of the discussion.

This utterance is executed on the condition that the low-resource language speaker does not speak for a certain period of time. The time at which a user sends a message is managed by the client side, so we implemented it on the client side. When a user sends a message, the server checks whether the user is speaking in an LRL. If three minutes have passed without any utterance from a low-resource speaker, the message “Let’s review the discussion so far” or “Let’s summarize the discussion once” will be randomly selected and sent under the user name SYSTEM MESSAGE to the chat room.

5 Experiment

5.1 Experimental Design

To study the effect of each strategy on the implemented system, we conducted a controlled experiment with a total of 19 subjects: five Indonesian speakers as LRL speakers, five Chinese speakers as HRL speakers, and 14 Japanese as HRL speakers. Each subject was either an undergraduate or graduate student. The subjects were divided into five groups and the effects of the facilitator agent’s utterances were examined through group discussions. Each group consisted of one LRL speaker, one Chinese speaker, and two Japanese speakers. The experiment was conducted over two days.

We prepared the following five experimental tasks (discussion themes) as shown in Table 2, and shuffled them after each discussion to attenuate the effect of the difficulty level of the tasks. The final goal of each task was to choose one answer from the given choices as a team and be able to explain the reasons why the choice was selected.

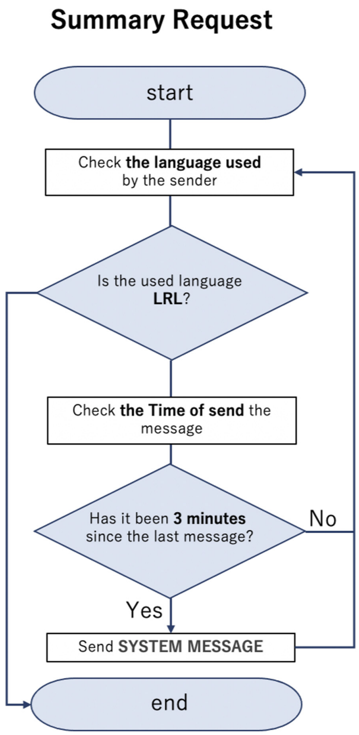

Summary Request
start
Check the language used by the sender
Is the used language LRL?
Check the Time of send the message
Has it been 3 minutes since the last message?
No
Yes
Send SYSTEM MESSAGE
end

Fig. 4. Flowchart of summary of the discussion request.

Table 2. Experimental tasks

Task No.	Experimental task	Given choice
1	If you could only take one thing to a desert island, which one would it be?	Lighter, knife, water, sleeping bag, fishing rod
2	If you were to adopt a new Olympic sport, which one would it be?	Bowling, tug of war, Frisbee, dodgeball, scuba Diving
3	Which of the following subjects do you consider most important?	Japanese, maths, science, social studies, English
4	If you were to be born again, which one would you want to be?	Bird, dog, cat, lion, dolphin
5	Which of the following points should you pay most attention to on a date?	Location, time, clothing, weather

After each discussion, the participants were asked to fill out a questionnaire for subjective evaluation. In addition, we also obtained chat log data for objective evaluation (Tables 4 and 5).

Table 3. Experimental set-up for each group

Group-time	Strategy	Method	Task No.	Language (Subject ID by language)
1-1	Request for a summary of the discussion	Bot	1	Low-resouce language (1) Chinese (1) Japanese (3) Japanese (4)
1–2	Request for rephrasing	Bot	2	Low-resouce language (1) Chinese (1) Japanese (2) Japanese (4)
1–3	Responses to the utterances from low-resource language speakers	Bot	3	Low-resouce language (1) Chinese (1) Japanese (1) Japanese (4)
2-1	Request for a summary of the discussion	Bot	3	Low-resouce language (2) Chinese (2) Japanese (1) Japanese (2)
2-2	Request for rephrasing	Bot	4	Low-resouce language (2) Chinese (2) Japanese (3) Japanese (2)
2–3	Responses to the utterances from low-resource language speakers	Bot	5	Low-resouce language (2) Chinese (2) Japanese (3) Japanese (2)
3-1	Request for a summary of the discussion	WOZ	1	Low-resouce language (3) Chinese (3) Japanese (7) Japanese (8)
3-2	Request for rephrasing	Bot	2	Low-resouce language (3) Chinese (3) Japanese (7) Japanese (5)
3-3	No communication	–	5	Low-resouce language (3) Chinese (3) Japanese (7) Japanese (3)
4-1	No facilitator agent	–	3	Low-resouce language (4) Chinese (4) Japanese (6) Japanese (3)
4-2	Request for rephrasing	WOZ	4	Low-resouce language (4) Chinese (4) Japanese (6) Japanese (8)
4-3	Responses positively to the utterances from low-resource language speakers	Bot	2	Low-resouce language (5) Chinese (5) Japanese (6) Japanese (9)
5-1	Responses positively to the utterances from low-resource language speakers	Bot	1	Low-resouce language (5) Chinese (5) Japanese (9) Japanese (5)
5-2	No facilitator agent	–	2	Low-resouce language (5) Chinese (5) Japanese (9) Japanese (3)
5-3	Responses to the utterances from low-resource language speakers	WOZ	3	Low-resouce language (5) Chinese (5) Japanese (5) Japanese (8)

6 Experiment Result

6.1 Response to Facilitator Agent Utterance

During the experiment, we evaluated the responses to the facilitator agent for the first two strategies since they are considered requests from the facilitator agent; the other two strategies are not directive.

For utterances requiring a summary of the discussion, the result is shown in Table 4. The effect of the action was lower in group 1. This is because the subjects did not respond to the system message and ignored it even though the facilitator agent sent it as programmed. In response to this, on the second day of the experiment, the system messages were changed from “Let’s review the discussion so far” and “Let’s summarize the discussion once” to “[Name], please tell me what you are talking about now”, “[Name], please review the discussion so far”, “[Name], what are you talking about now?”. The following is a list of the changes made to the previous section. The subjects who were named by other language speakers were more likely to respond to the system messages, and in fact, the subjects who were named by other language speakers responded to the system messages in Group 3 on the second day of the experiment. Group 2 was considered invalid because a significant result could not be obtained due to a malfunction of the system.

Table 4. Assessment of “Request for a summary of the discussion”.

Group	Facilitator agent utterance count	Effectiveness (Percentage of response)
1	6	17%
2	–	–
3	1	100%

As with the utterance requesting a summary of the discussion, the facilitator agent worked correctly here, but the subjects did not respond to the system message, so the effect of the action could not be discerned. On the second day of the experiment, the system message was changed from “Let’s summarize in simple words” and “Please rephrase briefly” to “Mr. [Name], could you rephrase what you just said in simple words?”, “Mr. [Name], could you please rephrase what you just said in simple words?”, “Can you please rephrase what you just said in simple words?”. However, the effectiveness of this strategy was zero. There was no response nor was any utterance rephrased.

6.2 Number of Utterances

Based on the objective evaluation, we analyze how the number of utterances of low-resource language speakers changed with each condition.

Table 5. Mean quantitative ratings for speakers of low-resource languages

Strategy\Question	Number of utterances by speakers of low-resource languages	Time taken to speak by speakers of low-resource languages	Number of opinions expressed by speakers of low-resource languages	Number of characters uttered by speakers of low-resource languages	Number of times a high resource language speaker has talked a low resource language speaker
Request for a summary of the discussion	4	3.14	4	333	0.67
Request for rephrasing	4.67	2.72	3	332	1.11
Responses to the utterances from low-resource language speakers	4.33	3.32	2	274	0.56
Responses positively to the utterances from low-resource language speakers	5.33	2.81	4	356	0.78
No facilitator agent	4.33	2.98	3.33	415	0.78

Table 6. Mean subjective ratings of low-resource language speakers

Strategy\Question	Comprehension of the content of the discussion	Comprehension of other participants' utterances	Ease of speaking up during discussions	The extent to which they feel they have been able to communicate their views to other participants	The extent to which the action by facilitator triggered utterance	The extend to which they felt able to participate in the discussion
Request for a summary of the discussion	4.5	4.5	4.5	4.5	4.5	5
Request for rephrasing	4.67	4.67	4	4.67	2.67	4.33
Responses to the utterances from low-resource language speakers	4.67	4.33	4.67	4.33	4	4.67
Responses positively to the utterances from low-resource language speakers	5	4.67	4.33	4.33	4	4.67
No facilitator agent	5	4.67	4.33	4.67	2	4.67

The following table summarizes the mean values of the results for each condition.

One-way ANOVA was performed on these values. However, no items were found to be significant.

6.3 Subjective Evaluation

Subjective evaluation was done with questionnaires. LRL speakers were asked to rate the following six items shown in Table 6, and speakers of other languages were asked to rate the following five items, as shown in Table 7, on a five-point scale. The numbers indicate the average of the subjects' answers.

One-way ANOVA on ranks was conducted on these results. The results showed that the LRL speakers had a P value of 0.0471 for the item "Did communication from the facilitator trigger your speech?", which was significant at the

Table 7. Mean subjective ratings of high-resource language speakers

Strategy\Question	Extent to which Indonesian participants were considered to have understood the discussions	Degree to which participants were considered to have felt uncomfortable talking	Degree to which they felt they were able to support others to participate in the discussion	Degree to which they were able to communicate their views clearly	Degree to which everyone felt included in the discussion
Request for a summary of the discussion	4.5	4.5	4.5	4.5	4.5
Request for rephrasing	4.67	4.67	4	4.67	2.67
Responses to the utterances from low-resource language speakers	4.67	4.33	4.67	4.33	4
Responses positively to the utterances from low-resource language speakers	5	4.67	4.33	4.33	4
No facilitator agent	5	4.67	4.33	4.67	2

5% significance level. As for the questionnaire for the other language speakers, there were no items that were significant at the same significance level. However, there was significance at the 10% significance level for the item "Did you feel that some people seemed to be difficult to talk to?".

7 Discussion and Future Direction

On the first day of the experiment, subjects did not respond to the system messages requesting them to summarize the discussion, but when the system messages were changed from a general message to a message that included a specific name, the subjects responded. This is thought to be because it became difficult to ignore the request. Therefore, we think that the facilitator agent should make utterances that give some sense of obligation to the subjects to respond.

Sending the paraphrase request utterance did not receive a good response. This may be due to the fact that the experimental task itself did not require long utterances and the difficulty level was not appropriate. Therefore, it is necessary to verify the effectiveness of this condition through group discussions with more difficult content or experiments outside group discussions. One potential issue with the restatement request is that the facilitator did not provide a clear direction for paraphrasing. Because the sender's new communications may still be complex, the translation quality might remain low.

Even though the questionnaire responses indicated that the LRL speakers thought that the facilitator triggered their utterances for some strategies, the number of utterances from the LRL showed no obvious difference.

Tasks requiring more effort to communicate might be more appropriate in our future experiment.

Since we could not measure the effects of utterance responding to LRL utterances, and positive responses to them, our future plan include adjusting the task and evaluation methods. In addition, more iterations of experiments should also be conducted to confirm the results presented here. The limitation of this study includes the design of the experiment task for each group of participants.

In the future, we plan to improve the facilitator agent so the users feel more obligated to respond, based on the ideas from existing research including making the conversation human-like as much as possible [3], starting from changing the user name of the agent to a human name and redesigning the response utterances. Another idea from previous research [10] is to embody our facilitator agent by using API for face and speech modalities.

8 Conclusion

In order to solve the communication problem caused by the limited accuracy of machine translation in multilingual communication, we defined utterances that were considered to be effective in activating communication by low-resource language speakers. We implemented these utterances in LangridChat, a multilingual chat system, and verified and analyzed their effectiveness through experiments with participants in group discussions.

We defined four strategies for the facilitator agent based on existing research. In order to facilitate the understanding of low-resource language speakers, we defined two types of utterances: one that requests a summary of the discussion, and the other that asks for a paraphrase. Furthermore, to facilitate low-resource language speakers' participation, we created two types of responses that provide positive responses to the opinions expressed by low-resource language speakers. The results of an experiment showed that there was significance at the 5% level in the subjective evaluation of "whether the communication from the facilitator triggered utterances". However, no significant difference was found in the results obtained from the objective evaluation.

Acknowledgements. This research is partially supported by a Grant-in-Aid for Scientific Research (A) (17H00759, 2017–2020), a Grant-in-Aid for Scientific Research (B) (21H03561, 2021–2024) and a Grant-in-Aid for Early-Career Scientists (21K17794, 2021–2024) from the Japan Society for the Promotion of Sciences (JSPS).

References

1. Chalidabhongse, J., Chinnan, W., Wechasaethnon, P., Tantisirithanakorn, A.: Intelligent facilitation agent for online web-based group discussion system. In: Hendtlass, T., Ali, M. (eds.) IEA/AIE 2002. LNCS (LNAI), vol. 2358, pp. 356–362. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-48035-8_35
2. Den, Y.: Adding information to dialogue. In: Koiso, H. (ed.) Lecture Japanese Corpus, vol. 3, pp. 101–130 (2015). (in Japanese)
3. Hwang, A.H.C., Won, A.S.: Ideabot: investigating social facilitation in human-machine team creativity. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–16 (2021)

4. Ishida, S., Inoue, K., Nakamura, S., Takanashi, K., Kawahara, T.: Listening dialogue system that generates listener responses for empathy expression and speech promotion. In: Materials of the Japanese Society for Artificial Intelligence Study Group Language/Speech Understanding and Dialogue Processing Study Group 82 times, p. 02. The Japanese Society for Artificial Intelligence (2018). (in Japanese)
5. Ishida, T., Murakami, Y., Lin, D., Nakaguchi, T., Otani, M.: Language service infrastructure on the web: the language grid. Computer **51**(6), 72–81 (2018)
6. Ito, T., Suzuki, S., Yamaguchi, N., Nishida, T., Hiraishi, K., Yoshino, K.: D-agree: crowd discussion support system based on automated facilitation agent. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 13614–13615 (2020)
7. Kim, S., Eun, J., Seering, J., Lee, J.: Moderator chatbot for deliberative discussion: effects of discussion structure and discussant facilitation. Proc. ACM Hum.-Comput. Interact. **5**(CSCW1), 1–26 (2021)
8. Kitaoka, N., Takeuchi, M., Nishimura, R., Nakagawa, S.: Response timing detection using prosodic and linguistic information for human-friendly spoken dialog systems. Inf. Media Technol. **1**(1), 296–304 (2006)
9. Pituxcoosuvarn, M., Ishida, T., Yamashita, N., Takasaki, T., Mori, Y.: Machine translation usage in a children's workshop. In: Egi, H., Yuizono, T., Baloian, N., Yoshino, T., Ichimura, S., Rodrigues, A. (eds.) CollabTech 2018. LNCS, vol. 11000, pp. 59–73. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-98743-9_5
10. Shamekhi, A., Liao, Q.V., Wang, D., Bellamy, R.K., Erickson, T.: Face value? exploring the effects of embodiment for a group facilitation agent. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–13 (2018)
11. Suzuki, S., et al.: Extraction of online discussion structures for automated facilitation agent. In: Ohsawa, Y., et al. (eds.) JSAI 2019. AISC, vol. 1128, pp. 150–161. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-39878-1_14
12. Yamashita, N., Inaba, R., Kuzuoka, H., Ishida, T.: Difficulties in establishing common ground in multiparty groups using machine translation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 679–688 (2009)

Towards Implementing Collaborative Learning in Remote Teaching Scenarios

Tommy Kubica(✉), Iris Braun, and Alexander Schill

Faculty of Computer Science, Technische Universität Dresden, Dresden, Germany
{tommy.kubica,iris.braun,alexander.schill}@tu-dresden.de

Abstract. The use of collaborative learning (CL) has been found to be beneficial for helping students to learn more effectively compared to learning on their own. While CL was implemented successfully in various classroom scenarios, it is used more rarely in remote settings. Especially during the CoViD-19 pandemic and its resulting distance learning, students were often used to learn alone, without experiencing the advantages of CL. This is caused by various reasons, ranging from the required technical know-how, hardware, and Internet connection, to the novelty and lack of experience with these functions. Moreover, the type of the remote setting, e.g., the number of participating students, is a challenge, as, similar to traditional face-to-face settings, the lecturer cannot follow every discussion. Given these challenges, the goal of this paper was to investigate how and, more importantly, how successfully CL can be implemented in remote settings. Therefore, the use of a novel technical approach to implement CL in four different remote teaching scenarios was investigated and both the usage itself as well as the opinions of lecturers and students were recorded. In this way, this paper intends to make a valuable contribution to the study of CL in remote settings.

Keywords: Collaborative learning · Peer feedback · Remote teaching

1 Introduction

Collaborative Learning (CL) is defined as a "situation in which two or more people learn or attempt to learn something together" [5]. In contrast to individual learning, the combination of resources and skills enables students to learn with or from each other. This active exchange of ideas in groups cannot only engage students in the learning process and thus, increase their interest but is also able to promote students' critical thinking [7,15,25].

While CL has been well studied and is already widely used in traditional school or university scenarios (cf. [3,18]), it faces several challenges in remote classrooms, which appear frequently as a result of the CoViD-19 pandemic [2]. First, it requires technical know-how for both lecturers and students, as well as suitable hardware and a stable Internet connection. Next, the novelty and lack of experience with technical functions to support CL are challenging, as best practices might not yet exist. Finally, challenges that arise in face-to-face

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 245–260, 2022.
https://doi.org/10.1007/978-3-031-20218-6_17

teaching, most likely also arise in the remote setting, e.g., the larger the number of students gets, the more difficult it is to create meaningful groups of students and the less aware the lecturer is of individual collaborations [19].

Taking these challenges into account, we attempt to answer the question of how to implement CL in different remote teaching scenarios and investigate both the associated opportunities and trade-offs. Therefore, a novel approach allowing the creation of customized teaching scenarios (including collaborative ones) will be implemented in four case studies. For each of them, the opinions of lecturers and students as well as log files are collected and analyzed afterward. The paper concludes with a discussion of the obtained results and provides an outlook on future contributions to be made.

2 Related Work

In the literature, there exist a variety of approaches to support both traditional face-to-face as well as remote teaching scenarios. According to [20], the following classroom interactions can be supported: *Teacher to individual student(s)*, *teacher to all students*, *student to teacher* and *student to student*. While early systems only allowed interactions to be initiated by the provision of polling questions (i.e., *Audience Response*) through the lecturer, the initiation of interactions by students was increasingly investigated in recent years.

On one hand, the use of *Backchannel* functionality was examined. A promising approach is presented by anonymous *Question and Answer* options that run in parallel to the ongoing lecture and allow students to create their own questions. Depending on the size of the lecture or the number of questions, they can either be answered directly by the lecturer, pre-selected by a moderator or rating options, or given to all students for discussion, e.g., as presented by the *Chatwall* [23]. While *Question and Answer* approaches have been known for a long time (cf. [1]) and were extended frequently (e.g., by marking helpful answers or selecting a type of question [4,10]), they are designed for collaborations between students of the entire classroom (or follow-up discussions between individual students) that run in the background of the lecture.

On the other hand, functions were investigated that allow integrating collaboration as an active part of the ongoing lecture, which is supported in different formats. The first approach is represented by *help requests* as an extension of *Audience Response* functions, which allow students to signal that they need help. These requests are sent to another student who has already answered the question [21]. However, these approaches are limited to the collaboration in pairs as well as to this particular use case. The collaboration in groups of students has so far been studied primarily in school scenarios with fixed technical tools, on which offline created groups of students work together (e.g., [16] or [6]). In contrast, [22] presents an approach that integrates the formation of groups in a digital learning environment and enables the assignment of questions to specific groups of students, in which interactions can be performed. In addition, [9] introduces collaboration as part of digital learning lessons. Students are given the opportunity to perform different tasks in a group using different roles to achieve a common goal. However, each of these approaches is targeted to small scenarios and fixed user accounts assigned to the students. Another promising approach is presented by [17] and describes a scalable, flexible CL approach that follows the *Pyramid* CL flow pattern. In iterative rounds of discussions and voting on student-created questions, it allows to reduce the list of questions to the most relevant ones. Nevertheless, this approach is limited to this specific scenario.

Defining individual CL scenarios is another research direction. For example, [24] presents an authoring tool allowing non-expert learning designers to combine best practices to describe CL situations. While this approach mainly focuses on supporting the design process, [8] describes an approach to author and execute so-called "Orchestration Graphs". This allows performing individual CL scenarios in a web-based application. Although the approach is very expressive, it is currently limited in the way the groups are formed. Similar problems are found in current live-stream systems, e.g., BigBlueButton[1] or Zoom[2], as well as Learning Management Systems, e.g., Moodle[3]. While these systems support collaboration, group formation is limited to manual or randomly formed groups.

In summary, there exist a variety of approaches to foster CL in remote settings. While current systems already allow the formation of groups of students as well as their collaboration, the meaningful formation of groups based on students' prior knowledge is not supported, nor are more advanced collaborative scenarios (e.g., *Jigsaw Classroom*). Consequently, in the next section, an approach will be presented that allows defining customized teaching scenarios, including expressive means for both collaboration and group formation.

3 An Approach to Support Customized Teaching Scenarios

This section will briefly describe the underlying approach that this paper relies on. Therefore, first, the general idea of the approach is described before specific collaborative components are further explained.

3.1 General Approach

In [13], the approach of an adaptable, collaborative learning environment called scenario-tailored Audience Response System (stARS) was presented that allows lecturers to configure the system's functionality to support their individual teaching strategy, including collaborative activities. In Fig. 1, each part of the concept is visualized and will be described in the following.

The foundation of the approach is a unified *metamodel*[4] that allows describing technology-enhanced lectures with their interactive activities as customized workflows. In addition to common workflow elements, such as a start node, end

[1] https://bigbluebutton.org/ – last access on August 21, 2022.

[2] https://zoom.us/ – last access on August 21, 2022.

[3] https://moodle.org/ – last access on August 21, 2022.

[4] A current version of the metamodel can be retrieved on https://stars-project.com/metamodel.pdf (last access on August 21, 2022).

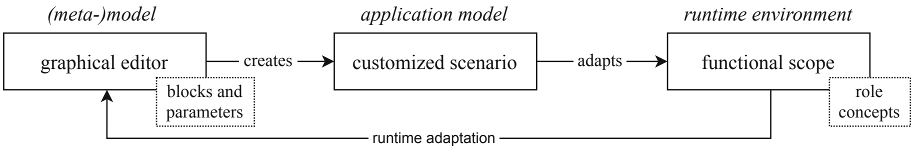
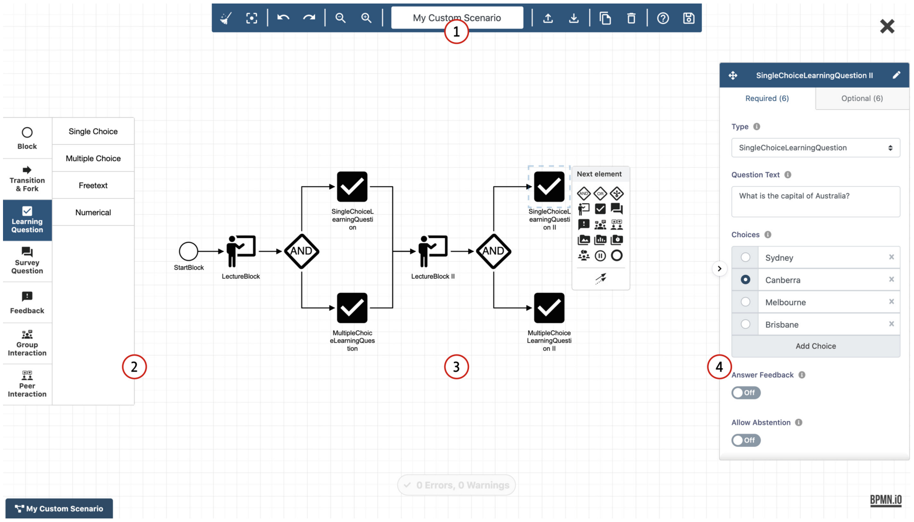


Fig. 1. The general concept of an adaptable, collaborative learning environment allowing lecturers to support customized teaching scenarios [11].

nodes, or forks, it defines a variety of functional blocks (i.e., interactive activities) that allow initiating interactions between the lecturer and the students or between students themselves. These include not only blocks for executing traditional *Audience Response* and *Backchannel* functionalities but also blocks to integrate expressive means of collaboration (such as the formation of groups of students and the subsequent discussion within these groups). Each of the defined blocks can be further customized by setting a variety of parameters, e.g., a *SingleChoiceLearningQuestion* can define a parameter *answerFeedback*, which, when being set to *true*, reveals the correct answer to the students after answering. The entire list of parameters can be retrieved from the *metamodel*.

In order to ease the creation of customized teaching scenarios, a graphical editor was developed allowing lecturers to create their individual workflows. The editor with its different components as well as an exemplary, unfinished workflow is displayed in Fig. 2.



SingleChoiceLearningQuestion II

Required (6) | Optional (6)

Type: SingleChoiceLearningQuestion

Question Text: What is the capital of Australia?

Choices:
○ Sydney ×
● Canberra ×
○ Melbourne ×
○ Brisbane ×
Add Choice

Answer Feedback: Off

Allow Abstention: Off

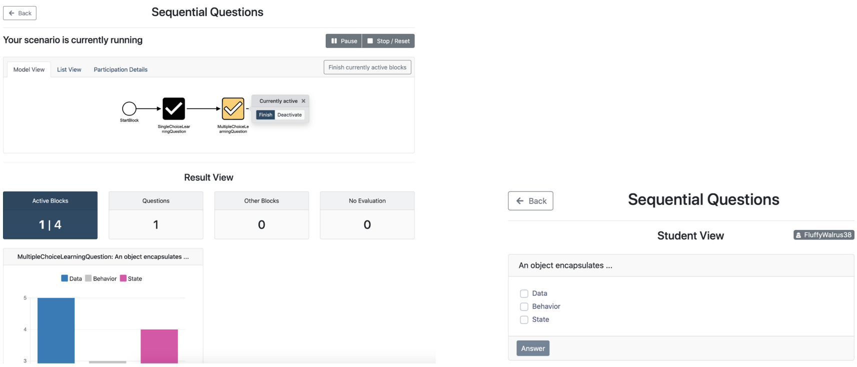
Fig. 2. The graphical editor allows lecturers to model customized workflows that represent their personal teaching strategies.

The editor consists of four different components:

(1) The *main menu* provides general controls (e.g., undo/redo, zooming, or saving) for the editor and allows specifying the name of the scenario.
(2) Next, the *element palette* lists all groups of activities with their specific types that can be drag and dropped into the scenario.
(3) The scenario itself is created within the *modeling canvas*, on which elements can be freely moved and connected.
(4) Clicking on a certain element in the *modeling canvas* opens the *properties panel*, in which both required and optional parameters can be defined.

The workflows created in this way are stored on a backend server, from which they can be started by the lecturer. When starting a scenario, a container is created on one of the cloud servers and initialized with the respective workflow. The backend server manages the information about which instances are running on which cloud servers, how they can be contacted, and which users can access them. It is also responsible for authenticating the users and issues tokens used to authenticate against the individual instances in all subsequent interactions.

In contrast to the backend server, the instances that run on the cloud servers contain the actual functionality. Depending on the active functional blocks, users can interact with the instance and, for example, answer questions (cf. Fig. 3b). Lecturers can view the results and proceed within the scenario by finishing functional blocks (cf. Fig. 3a). The interaction between the clients and the instances takes place largely via a REST API. Furthermore, WebSocket connections to each student are established in order to inform them about changes in the workflow and to enable bi-directional real-time communication, as it takes place, for example, in the *GroupChat* with the group members.



(a) The lecture view with the options to finish currently active functional blocks and to view the results.

(b) The student view, in which an interaction with the currently active functional blocks is possible.

Fig. 3. An exemplary scenario visualizing both the lecturer and the student view.

The concept also includes an implementation of the *role concept* that allows adding dynamic functionalities, which cannot be expressed by the *metamodel* itself, more intuitively (cf. [14]). One example is the assignment of students into specific groups, which is done at runtime and cannot be specified before. In addition, the means of *runtime adaptation* are implemented, allowing the extension of currently running scenarios. Possible reasons could be incoming results or the realization of student-centered approaches (cf. Sect. 4.5).

The presented approach allows supporting a variety of teaching scenarios while being mostly limited to synchronously held scenarios (either face-to-face or remote ones). Four specific case studies as well as an outlook on further scenarios are presented in Sect. 4.

3.2 Collaborative Components

The *metamodel* includes a variety of collaborative components to execute both *group* and *peer interactions*. In the following, each of these components will be described in further detail.

Before interactions in groups can be conducted, groups of students must be formed. Therefore, a *GroupBuilder* is used and later referred in all types of *group interactions*. This *GroupBuilder* defines a variety of parameters: Either a *groupSize* or a *numberOfGroups* has to be set to determine either the size of groups or the number of groups to be created. Furthermore, a *buildSchema* has to be selected, which is used to form groups of students. As stARS supports a variety of further functional blocks, such as *learning questions*, those results can be used for creating meaningful groups of students. The following build schemes are supported: *random*, *bestToWorst*, *similar*, *sameAnswer*, *differentAnswer*, *groupShuffle* and *groupMerge* (cf. [11]). The *GroupBuilder* is followed by different *group interactions*:

- A *GroupChat* allows group members to exchange textually,
- a *GroupAudioVideoChat* allows communicating using audio and video,
- *PresentGroupAnswers* lists the previously given answers of all group members to a specified question, and
- a *GroupVoting* is used to allow selecting a common group answer.

Another type of collaborative components is presented by *peer interactions*, e.g., to execute *peer feedback*. Similar to *group interactions*, a *PeerBuilder* is required to create assignments for students. Using the parameter *numberOfAssignments* allows setting multiple feedbacks to be created and collected. Moreover, an arbitrary number of *SurveyQuestions* has to be selected, on which the feedback should be collected. An example would be the upload of a paper that should be rated by another student. Furthermore, one of the following build schemes can be selected: *random*, *bestToWorst*, *similar*, *sameAnswer*, *differentAnswer*, *sameGroup* and *differentGroup* (cf. [11]). The *PeerBuilder* is followed by different *peer interactions*:

- *PresentPeerAnswers* displays the given answer of a student to one or multiple assigned students.
- *SurveyQuestions* can be used and referred to a *PeerBuilder* to provide the actual feedback. For instance, a *FileUploadSurveyQuestion* could be used to allow uploading a file containing the review.
- *PresentPeerFeedback* is used to display the provided feedback, e.g., the uploaded review file of the feedback provider.
- Finally, a *PeerChat* allows initiating textual discussions between peers.

Contrary to *group interactions*, in *peer interactions*, two students do not necessarily have to be assigned to each other. For example, having three students *s1*, *s2* and *s3*, *s1* could provide feedback to *s2*, *s2* to *s3* and *s3* to *s1*.

4 Case Studies to Integrate Collaborative Learning in Remote Settings

In this section, four case studies of remote teaching scenarios will be presented, in which stARS was implemented. Even though each of these scenarios includes at least one collaborative activity, it is important to note that the complete teaching scenario was modeled. For each of the case studies, the results of the execution as well as both lecturers' and students' opinions will be discussed. This is done by the following means of evaluation: First, the execution of the scenarios was observed and log files were analyzed. Second, lecturers' opinions were recorded by questionnaires and interviews, if required. Finally, each scenario included two questions at its end to retrieve students' opinions as well.

4.1 First Case Study

The first case study was conducted on June 14, 2021 in a remote lecture of the BA Dresden[5] by Dr. Marius Feldmann. The goal of this lecture is to teach the basics of Linux to students in the second semester of the study courses *Information Technology and Media Computer Science*. Therefore, it includes both a regular part (i.e., a presentation held by the lecturer) and a practical part. Using stARS, Dr. Feldmann intended, on one side, to improve the lecture's structuring and, on the other side, to increase the interaction between the lecturer and the students and also between students themselves. Therefore, the lecture was represented by a workflow including different means of interaction, as visualized in Fig. 4.

The regular parts of the lecture were supported by two rounds of questions, in which students could first recapitulate their prior knowledge and select a topic of interest (*SingleChoiceSurveyQuestion*), before checking their gained knowledge using four distinct *LearningQuestions* after the first presentation was held. Afterward, the second presentation was held and supported by an advanced collaborative scenario. The latter started with the description of the topic of

[5] https://ba-dresden.de/ – last access on August 21, 2022.

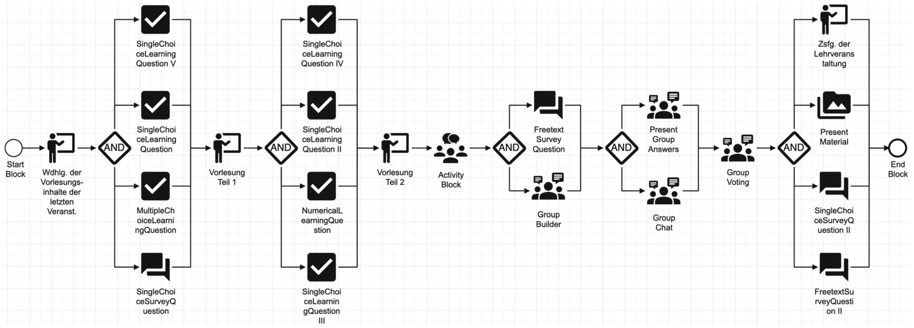

Start Block
Wdhlg. der Vorlesungs-inhalte der letzten Veranst.
AND
SingleChoiceLearning Question V
SingleChoiceLearning Question
MultipleChoiceLearningQuestion
SingleChoiceSurveyQuestion
Vorlesung Teil 1
AND
SingleChoiceLearning Question IV
SingleChoiceLearning Question II
NumericalLearningQuestion
SingleChoiceLearningQuestion III
Vorlesung Teil 2
Activity Block
AND
Freetext Survey Question
Group Builder
AND
Present Group Answers
Group Chat
Group Voting
AND
Zsfg. der Lehrveranstaltung
Present Material
SingleChoiceSurveyQuestion II
FreetextSurveyQuestion II
End Block

Fig. 4. The scenario of the first case study conducted by Dr. Marius Feldmann, which includes two rounds of questions as well as an advanced collaboration.

the practical part (*ActivityBlock*), followed by the upload of an individual solution (*FreetextSurveyQuestion*) and the acknowledgment of joining the collaborative task (*GroupBuilder*), in which students should collaborate in groups. Contrary to related approaches, stARS allowed forming meaningful groups of students by defining the *buildSchema bestToWorst* and referring it to the previously answered *LearningQuestions*. In addition, a *groupSize* of 4 was set, targeting to have at least one student in each group having prior knowledge of the topic. The collaboration within these groups was the next part of the scenario and included both the visualization of group members' given answers (*PresentGroupAnswers*) as well as a textual group discussion (*GroupChat*). After a certain amount of time, the lecturer could unlock the voting for a common group answer (*GroupVoting*) that helps to reduce the number of responses to the number of formed groups, which can then be discussed in more detail. The scenario was concluded with a summary and two questions for evaluation purposes.

The usage of stARS was announced in advance of the lecture. Additionally, an explanation of both the reason for using the tool as well as the structure of the lecture was given. Dr. Feldmann estimates that about 90% utilized the system, whose usage took around 20 to 30 min. While both rounds of questions could be carried out without problems, technical issues were encountered in the practical part. The first problem was the dependency of both the group discussion and group voting on the responses to the *FreetextSurveyQuestion*. As only 16 of 42 students (who acknowledged joining the group collaboration) answered and the groups were formed based on the results of the previous rounds of questions (the *buildSchema bestToWorst* was defined), groups existed, in which none of the group members gave an answer beforehand. Thus, also no answer could be discussed, resulting in empty or incomplete group answers. Another problem occurred in the group formation. Instead of forming eight groups of four students and two groups of five students, ten groups of four students were formed and one group including only two students. Later, the implementation was adjusted.

Even though the case study experienced some problems, Dr. Feldmann was satisfied with the “simple usage” and “the possibility of structuring the lecture and checking the level of knowledge of the students more easily”. Furthermore, he was willing to use it again in another lecture (cf. Sect. 4.4). The students rated the approach quite positively, as well. However, some encountered the previously described technical problems, which could also be recognized in the results: Having 39 students answering (with two abstentions), only five agreed that the system could be used analogously to similar systems, 18 partly agreed, 13 did neither agree nor disagree and one did partly disagree. Using a rating from 0 (disagree) to 4 (agree), an average rating of 2.73 can be recorded, indicating that students partly agree with the statement. Furthermore, from 16 textual answers, seven stated that they experienced problems with the *GroupVoting*. However, other students praised the benefits of the approach: Five students liked the questioning function and three the idea of virtual group discussions.

To further evaluate the functionality, the textual messages exchanged in the *GroupChat* were analyzed. All eleven groups used the textual chat, while the number of messages varied from six to 20, with a total of 114 messages exchanged. It could be recognized that most of the group chats (nine out of eleven) started with “hello” or “test”. This was caused by the state of the implementation that students were not shown how many students they were in a group with. However, ten out of eleven groups discussed the topic of interest. Nevertheless, only three of those actually discussed about finding a concrete answer. One surprising fact, however, is the low number of spam. Although students could enter as many and as long messages as they wanted, and would have been protected by their anonymity, nobody made use of it. Also, no abusive messages were sent.

4.2 Second Case Study

The second case study was conducted between June 18 and June 28, 2021 in a remote seminar of the TU Dresden[6] by Dr. Iris Braun. The seminar is addressed to students of computer science and runs over a complete semester. It works analogously to a conference setting and includes the creation of a paper about a scientific topic, a *peer feedback* to rate each other’s paper as well as a presentation. stARS was utilized to support the *peer feedback* scenario. Therefore, a three-step scenario was created, as visualized in Fig. 5.

In the first phase, the students were able to submit their papers and had to acknowledge to join the peer task (*PeerBuilder*), which was described textually using the *PresentMaterial*-block. Afterward, the assignments (*numberOfAssignments* was defined to 1) were made and each student was presented another paper (*PresentPeerAnswers*), for which a review had to be created and uploaded. Finally, after all reviews were submitted and the next stage was unlocked, the students received their feedback (*PresentPeerFeedback*) and could discuss it with the creator of the feedback (*PeerChat*). Furthermore, similarly to the first case study, two questions for evaluation purposes were added.

[6] https://tu-dresden.de/ – last access on August 21, 2022.

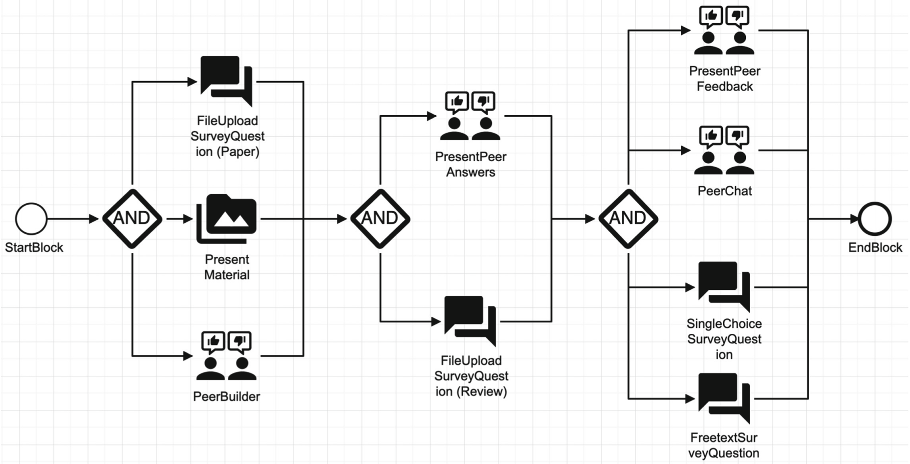


Fig. 5. The scenario of the second case study conducted by Dr. Iris Braun, which includes three steps to represent *peer feedback*.

The usage of the system was announced in the Learning Management System used[7] and a detailed instruction of the scenario was provided to the students. In total, 14 students participated and both uploaded their papers and acknowledged joining the peer task. However, after finishing the first step (and after the assignments were already made), one student had to leave the seminar as his/her submitted paper did not met the requirements to pass. Thus, the assignment of the student had to be executed by another student or the supervisor itself. We decided to change the student's password (logins only consist of a pseudonym and a password), so the supervisor could upload the missing feedback and potential spam was avoided. For future scenarios, an assignment management was implemented, allowing to manually add or remove assignments.

Besides the challenge of one student dropping out, no technical errors were encountered. Thus, Dr. Braun praised both the automatic assignment of peers and the opportunity to exchange with the feedback provider at the end of the scenario. Moreover, she made two proposals for improving the usage, namely, the addition of further textual descriptions and the access of multiple lecturers (i.e., supervisors) to the submissions of the students. Such improvements were part of two further runs of the scenario at the beginning and mid of 2022.

From 13 students participating in the case study, only four gave feedback on stARS, of which two fully agreed that the system could be used analogously to related systems, while two partly agreed. Due to the low percentage of students answering, this can only be seen as an impression rather than representing the opinion of the majority of students. In the textual responses, two students rated the anonymity of the system positively, while another two would propose to add more textual descriptions to the blocks. In the *PeerChat*, a total of eight

[7] https://bildungsportal.sachsen.de/opal – last access on August 21, 2022.

messages were exchanged. While two students exchanged their opinions about the feedback, another student tried, but his/her feedback receiver did not reply.

As the majority of participants did not reply to the questions in the third step, we asked them about their opinion after completing their presentations. In general, the students stated that they could use the system as described. Some of them expected further functions to be supported, such as using different questions to rate the submissions – this is already supported in stARS but was not included in the scenario. Other students argued that similar functions could be realized using the existing Learning Management System. However, it does not allow randomly assigning students' submissions to each other, does not provide means of discussion and finally, cannot guarantee the anonymity of the students.

4.3 Third Case Study

The third case study was conducted on June 23, 2021 in a virtual block seminar on educational-psychological interventions of the TU Berlin[8] by Dr. Felix Kapp. The seminar runs four days having four to six teaching units (i.e. 45 minutes) each. Dr. Kapp used several tools to improve the interaction, e.g., a traditional Audience Response System to assess students' gained knowledge. stARS was used on the last day and targeted to execute a task ("create an intervention of your own") that involves students receiving and giving *peer feedback*. Therefore, a scenario taking two teaching units was modeled as shown in Fig. 6.

At the beginning of the scenario, stARS was introduced to the students and the general task was described (*LectureBlock*). Afterward, the students should select a topic for their intervention (*SingleChoiceSurveyQuestion* having four options to choose from). Depending on the selected option, they were shown specific material in the next stage (*PresentMaterial (I –IV)*), which was realized using the *filter* parameter. With this material, they had to create and upload their intervention (*FileUploadSurveyQuestion*) in a defined time frame of 20 minutes (*PresentCountdown*). Furthermore, a combination of a *GroupBuilder* and *PeerBuilder* was modeled to implement the intention of the lecturer. The goal was that two students who choose the same topic for their intervention should give feedback to each other and discuss it afterward. As the *PeerBuilder* does not guarantee that the same students give feedback to each other, a *GroupBuilder* was used to form groups of two students using the *buildSchema sameAnswer*, followed by the *PeerBuilder* referring to this *GroupBuilder*. In the next step, the students were displayed their assigned submission (*PresentPeerAnswers*) and were asked to rate it on different dimensions using four *FreetextSurveyQuestions*. Additionally, they were displayed a countdown of 15 minutes and a description of the task. Afterward, the students received their feedback (*PresentPeerFeedback (individual)*) and had a time frame of five minutes to review it, before meeting their peer in an audio-video chat. Finally, Dr. Kapp concluded the scenario and the students were presented with the two questions for evaluation purposes.

[8] https://tu.berlin/ – last access on August 21, 2022.

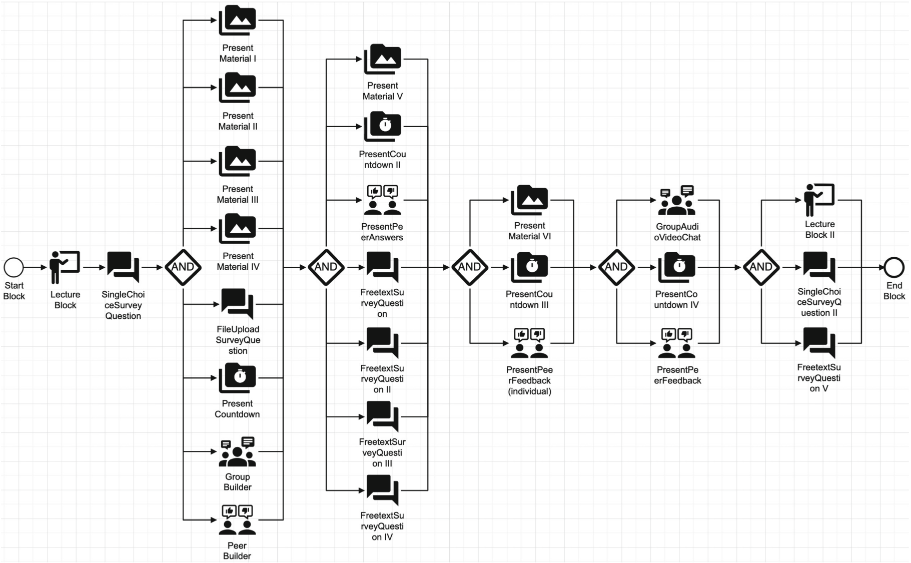


Fig. 6. The scenario of the third case study conducted by Dr. Felix Kapp, which realizes an advanced *peer feedback*.

Eleven out of 15 students that were enrolled in the seminar were present on the fourth day of the seminar, while ten of them participated in the presented scenario. Despite one student having problems with his/her computer due to forced updates of the operation system, which sometimes caused other students to wait, the system could be used without errors, resulting in five groups with two students each being created. Dr. Kapp praised stARS as an approach "to create a very organized *Peer Feedback* situation online, which is not yet supported by existing systems." In an interview, he pointed out the importance of such scenarios, which were essentially in traditional settings but could not be fully transferred to remote settings. Moreover, he emphasized the ability of the system to make decisions for problems that he did not consider before, e.g., assigning students, if their selected topics do not match. Even though the scenario can be reused 1:1, the representation is not trivial to understand and most likely requires an expert to recreate. However, it was able to illustrate the expressiveness of the approach to model a scenario that was not explicitly considered in advance.

The opinions of the students differentiate. While five students did fully agree that the system could be used analogously to related ones, two did partly agree, two did neither agree, nor disagree and one did partly disagree. Using a rating from 0 (disagree) to 4 (agree), an average rating of 3.1 results, indicating a partial agreement among the students. Textual responses were submitted by nine out of ten students: Even though one student had problems uploading his/her submission, positive feedback dominates: Three students stated that everything worked as expected, two found the approach intuitive to use and one student especially praised the novelty of the approach and stated that he/she never saw a similar approach before. Furthermore, another student experienced that the usage "felt like real partner work". Using the log files, the correctness of the built groups could be verified: With a distribution of 5, 2, 1 and 2 among the selected topics, one mixed group was created.

4.4 Fourth Case Study

The fourth case study was conducted on July 22, 2021 as a part of a virtual exam consultation on internet and web applications of the TU Dresden (see footnote 6) by Dr. Marius Feldmann. The goal was to structure the scenario more efficiently and recapitulate the topic of *Kademlia*, which was a subject relevant to the exam. Moreover, during the scenario, a student thesis was presented, which was also represented within stARS. The resulting scenario is visualized in Fig. 7.

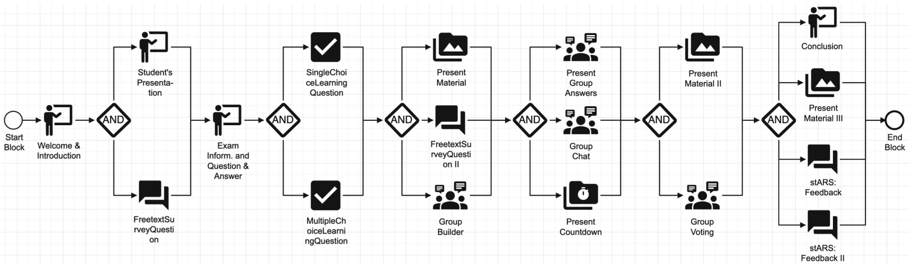


Fig. 7. The scenario of the fourth case study conducted by Dr. Marius Feldmann, including a student presentation, a round of questions and a group collaboration.

The scenario started with an introduction, in which the workflow's structure was explained. Afterward, the student shortly presented his/her master thesis and the students were asked to submit their email addresses if they were willing to help evaluate it. Next, the exam information was presented before students could ask their own questions using audio or chat in the corresponding Zoom session. After all questions were clarified, the interactive part started. Therefore, the students were first asked to check their gained knowledge on *Kademlia* using two learning questions. After discussing the results, the students should describe the functioning of *Kademlia* in no more than 500 characters (*FreetextSurveyQuestion II*). Therefore, both a reference to the slides and a paper were presented. Additionally, a *GroupBuilder* was added, in which students had to acknowledge whether they are willing to join the collaborative task. Similar to the first case study, the *buildSchema bestToWorst* was specified with a *groupSize* of five students. As soon as this step was finished and the groups were formed, the students were presented their group member's given answer and could discuss textually. Therefore, a fixed time frame of five minutes was defined. In the next step, each group had to select a common group answer by voting on the previously submitted answers. If no common group answer was found, a randomly selected moderator of the group could input a new answer. After all group answers were found and discussed by Dr. Feldmann, the lecture was concluded and the students were asked to answer two questions for evaluation purposes.

The usage of stARS was not announced beforehand, but introduced at the beginning of the scenario. While 60 students were present at the beginning, half of those left after the questions about the exam were clarified. However, 32 answers were recorded for the *SingleChoiceLearningQuestion* and 33 for the *MultipleChoiceLearningQuestion*. Nevertheless, this further decreased within the next step, in which only 13 students submitted a textual answer, while 16 joined the group interaction. As a result, three groups were formed, one group having six students. Those 16 students actively used the system. Moreover, it could be recognized that another seven students passively joined the task, i.e., even if they did not participate in the group interaction, they answered the questions at the end. As no technical errors were encountered, we believe that the students were preparing for their exams and did not recognized the importance of the task.

However, Dr. Feldmann was satisfied with the usage of the system and praised the opportunity of using collaboration to actively repeat an important topic (i.e., *Kademlia*). Moreover, the reduction of 13 textual answers to three answers allowed for discussing those in more detail. Positive feedback is also reflected in the students' answers. When asked whether the system could be used similar to related systems without limitations, ten students agreed, nine did partly agree and one student each did rather not agree and disagree. Another two students did abstain from answering. Using a rating from 0 (disagree) to 4 (agree), an average rating of 3.24 can be recorded, indicating a partial agreement ranging toward agreement. Similar results are provided by the textual responses. One student stated that he/she found the model easy to understand. Another student liked the anonymity and one student praised that using stARS is "more interactive rather than just watching videos". However, two students also emphasized the necessity of a second screen to use the system more efficiently.

Using the logs, we could verify the correctness of the formed groups, which also holds for a similar knowledge distribution among its members (according to the *buildSchema bestToWorst*). In the three groups, a total of 22 textual messages were exchanged, each chat having between four and nine messages. Only one of the chats started with "Hi," as contrary to the first case study, the students were presented with the number of students inside their group. While one group was unsure about how selecting a group's common answer (which was part of the next step of the scenario), the *GroupVoting* itself worked without problems. This allowed reducing the number of 13 textual responses to three.

4.5 Applicability to Further Scenarios

Even though the previously presented case studies could motivate the importance of using CL in remote settings, the usage of stARS is not limited to these four scenarios. Instead, it is possible to model individual scenarios as well as adjust or extend existing ones – both having the same goal to support lecturers' individual teaching strategies. In order to visualize the expressiveness of the approach and motivate future use cases, several popular didactic scenarios were modeled prototypically, including *Peer Instruction*, *Jigsaw Classroom*, *Think-Pair-Share*, *Learning Stations* as well as *Learners-as-Designers*. These models were then rated by 20 lecturers regarding their understandability (cf. [12]). Similar to the results of Sect. 4.3, the observation can generally be summarized as follows: The more complex the scenario becomes, the less comprehensible it is. Nevertheless, the results also indicate that despite its complexity, experienced lecturers were still able to understand highly complex scenarios, such as *Learners-as-Designers* being one example. This specific scenario also highlights the flexibility of stARS, as besides static workflows, it includes concepts of *runtime adaptation*, allowing to adjust running scenarios, e.g., to react to results.

5 Conclusions

Within this paper, we described the usage of a novel approach to foster CL in remote teaching scenarios. Therefore, the underlying approach was briefly presented, which allows lecturers to configure the system's functionality to their targeted teaching scenarios. In addition to traditional *Audience Response* and *Backchannel* functionality, it includes components to integrate both *group* and *peer interactions* in the lecture (cf. Sect. 3.2). By describing four case studies, we were able to motivate the opportunities when integrating CL in remote scenarios. Not only do lecturers want to implement CL scenarios in remote settings, but students also recognize the opportunities that were not possible in individual learning as caused by the CoViD-19 pandemic. However, also the challenges of using novel approaches got obvious. Due to the novelty of the functionality as well as the inexperience of the users with such functions, several technical problems were detected that required improvements. Moreover, even though most of the functions described in Sect. 3.2 were included in the case studies, there exist a variety of scenarios that have not yet been validated (cf. Sect. 4.5). Thus, further case studies (also for larger scenarios) have to be conducted to provide another step towards implementing CL in the large majority of lectures.

References

1. Aagard, H., Bowen, K., Olesova, L.: Hotseat: opening the backchannel in large lectures. Educause Q. **33**(3), 2 (2010)
2. Aini, Q., Budiarto, M., Putra, P.O.H., Rahardja, U.: Exploring e-learning challenges during the global COVID-19 pandemic: a review. J. Sistem Informasi **16**(2), 57–65 (2020)
3. Bruffee, K.A.: Collaborative learning: higher education, interdependence, and the authority of knowledge. In: ERIC (1999)
4. Bry, F., Pohl, A.Y.S.: Large class teaching with backstage. J. Appl. Res. Higher Educ. (2017)

5. Dillenbourg, P.: What do you mean by collaborative learning? (1999)
6. Dragon, T., et al.: Metafora: a web-based platform for learning to learn together in science and mathematics. IEEE Trans. Learn. Technol. **6**(3), 197–207 (2013)
7. Gokhale, A.A.: Collaborative learning enhances critical thinking. J. Technol. Educ. **7**(1) (1995)
8. Håklev, S., Faucon, L., Olsen, J., Dillenbourg, P.: Frog, a tool to author and run orchestration graphs: affordances and tensions (2019)
9. Jagušt, T., Botički, I.: Mobile learning system for enabling collaborative and adaptive pedagogies with modular digital learning contents. J. Comput. Educ. **6**(3), 335–362 (2019). https://doi.org/10.1007/s40692-019-00139-3
10. Jiranantanagorn, P., Shen, H., Goodwin, R., Teoh, K.K.: Classense: a mobile digital backchannel system for monitoring class morale. Int. J. Learn. Teach. **1**(2), 161–167 (2015)
11. Kubica, T.: Supporting Lecturers in Properly Using Digital Learning Environments. Ph.D. thesis, TU Dresden (2022)
12. Kubica, T., Damnik, G., Braun, I., Peine, R., Schill, A.: Lecturers' perceptions of supporting digital teaching scenarios by an adaptable learning environment. In: EDULEARN. IATED (2021)
13. Kubica, T., Shmelkin, I., Peine, R., Roszko, L., Schill, A.: stARS: proposing an adaptable collaborative learning environment to support communication in the classroom. In: CSEDU. SCITEPRESS (2020)
14. Kühn, T.: A family of role-based languages. Ph.D. thesis, TU Dresden (2017)
15. Laal, M., Ghodsi, S.M.: Benefits of collaborative learning. Procedia-Soc. Behav. Sci. **31**, 486–490 (2012)
16. Lingnau, A., Kuhn, M., Harrer, A., Hofmann, D., Fendrich, M., Hoppe, H.U.: Enriching traditional classroom scenarios by seamless integration of interactive media. In: ICALT, pp. 135–139. IEEE (2003)
17. Manathunga, K., Hernández-Leo, D.: Authoring and enactment of mobile pyramid-based collaborative learning activities. Br. J. Edu. Technol. **49**(2), 262–275 (2018)
18. Meyers, C., Jones, T.B.: Promoting Active Learning. Strategies for the College Classroom, ERIC (1993)
19. Rannastu-Avalos, M., Siiman, L.A.: Challenges for distance learning and online collaboration in the time of COVID-19: interviews with science teachers. In: Nolte, A., Alvarez, C., Hishiyama, R., Chounta, I.-A., Rodríguez-Triana, M.J., Inoue, T. (eds.) CollabTech 2020. LNCS, vol. 12324, pp. 128–142. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58157-2_9
20. Robbins, S.: Beyond clickers: using ClassQue for multidimensional electronic classroom interaction. In: SIGCSE, pp. 661–666 (2011)
21. Sampaio, B., Morgado, C., Barbosa, F.: Building collaborative quizzes. In: Koli Calling, pp. 153–159 (2013)
22. Seralidou, E., Douligeris, C., Gralista, C.: Eduapp: a collaborative application for mobile devices to support the educational process in greek secondary education. In: EDUCON, pp. 189–198. IEEE (2019)
23. Vetterick, J., Garbe, M., Cap, C.H.: Tweedback: a live feedback system for large audiences. In: CSEDU, pp. 194–198 (2013)
24. Villasclaras-Fernández, E., Hernández-Leo, D., Asensio-Pérez, J.I., Dimitriadis, Y.: Web collage: an implementation of support for assessment design in cscl macro-scripts. Comput. Educ. **67**, 79–97 (2013)
25. Warsah, I., Morganna, R., Uyun, M., Afandi, M., et al.: The impact of collaborative learning on learners' critical thinking skills. Int. J. Instr. **14**(2), 443–460 (2021)

Scaffolding of Intuitionist Ethical Reasoning with Groupware: Do Students' Stances Change in Different Countries?

Claudio Álvarez[1,2](✉), Gustavo Zurita[3], Antonio Farías[3], César Collazos[4], Juan Manuel González-Calleros[5], Manuel Yunga[6], and Álvaro Pezoa[7]

[1] Facultad de Ingeniería y Ciencias Aplicadas, Universidad de los Andes, Santiago, Chile
calvarez@uandes.cl

[2] Centro de Investigación en Educación, Universidad de los Andes, Santiago, Chile

[3] Facultad de Economía y Negocios, Universidad de Chile, Santiago, Chile
gzurita@fen.uchile.cl

[4] Departamento de Sistemas, Universidad del Cauca, Popayán, Colombia

[5] Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, Puebla, Mexico

[6] Universidad Técnica Particular de Loja, Loja, Ecuador

[7] Facultad de Filosofía y Humanidades, Universidad de los Andes, Santiago, Chile

Abstract. Ethics education is essential in business and STEM curricula. For decades up to the present day, a rationalist conception of ethics has been highly influential in its pedagogy. However, in the past twenty years, developments in moral psychology and neuroscience support that moral thought and deliberation are guided by a dual process initiated by intuition. In this research, we present the design of a scaffolding for ethics teaching based on groupware with individual and collaborative activities, informed by the Social Intuitionist Model (SIM) and Moral Foundations Theory (MFT). We conducted an exploratory study based on an original case about academic ethics, involving student samples in five higher education institutions in four countries (N = 249). Results indicate that ethical reasoning, initially guided by intuition, can be influenced by facts, reflective questions, and social interaction. Students regardless of their initial intuitive stance about an eliciting situation, and their moral sensitivity according to MFT, changed their final decision on the situation significantly by the end of the intervention according to a Wilcoxon signed rank test ($p < 0.001$). About 30% of the sample swung to a stance opposite to what they decided at the outset of the intervention. A question that stems from this research is how students' ethical thinking could be scaffolded and oriented, both pedagogically and technology-wise, towards identifying and standing for solutions to ethical dilemmas that are based on virtue and bring greater good.

Keywords: Ethics teaching · Groupware · Social intuitionist model

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 261–278, 2022.
https://doi.org/10.1007/978-3-031-20218-6_18

1 Introduction

Ethics guides people's behavior and allows them to understand, analyze and distinguish between what is right and what is wrong, what is good and what is bad, what is admirable, and what is deplorable, all based on morality [1, 2]. Ethics is considered as an active process rather than a static one, where people support their beliefs and affirmations initially through an intuitive justification, and later rationally [3, 4], through their moral intuitions, logical arguments, the context, rules established, feelings and emotions [5].

Today there is a growing need for ethics education in university contexts [6, 7]. Traditionally, ethics teaching in higher education has been of the 'reactive' type, informing students about legal and regulatory issues or imparting knowledge of ethical codes of conduct, law, and ethical errors [2, 8]; lacking a 'proactive' attitude in the decisions of the students, so that they adopt ethical postures and socially responsible thoughts in the performance of their future work environments. This encourages higher education institutions to innovate in the pedagogy of an ethical education focused on students, 'proactive' and supported by methodologies that adjust to the current requirements and characteristics of students of recent generations, such as millennials and Z [5], using active learning [9] and case-based learning [10] methodologies. In [6, 10], the need to develop learning experiences that involve situations where students reflect not only on their own emotions but also on the emotions of others and can develop greater empathy is highlighted. This has been supported by training criteria and accreditation requirements proposed by institutions such as the AACSB, ABET, AAA, IEEE, etc. [11, 12].

The bibliographic review of [7] regarding ethical teaching, identified that between 1991 and 2000, one of the main challenges was how it should be incorporated into the educational curriculum of higher education in a way that improves **moral judgments** and reasoning criticism of students facing ethical dilemmas. Likewise, between 2011 and 2020, Poje and Zaman [7] identified the need to teach ethics in an integral way to the academic curriculum, the use of improved ethics frameworks, and the development of professional values [7]. These authors report that there is still no consensus on how ethics should be taught, recommending the use of innovative methods, such as **active learning** [9], **case-based learning** [6, 10], thematic approach, virtue ethics, giving voices to values, role-playing, among others, over traditional methods. Poje and Zaman [7] propose future research in ethical teaching, the design of new online teaching methods, and compare their results against traditional methods.

Moral psychology has long been dominated by rationalist models of moral judgment [5], where moral judgment is reached mainly through processes of reasoning and reflection [13]. According to this approach, moral emotions can be inputs to the reasoning process but are not seen as direct causes of moral judgments.

On the other hand, intuitionism refers to the fact that there are moral truths, where people do not apply rational or reflective processes but rather processes more similar to perception, in which one "simply sees without discussion that they are and should be true". Intuitionistic approaches in moral psychology, by extension, say that moral intuitions, including moral emotions, come first and directly cause moral judgments [3]. This has been reflected in the **Moral Foundation Theory (MFT)** [5, 14] and the **Social Intuitionist Model (SIM)** [3, 15]. The MFT proposes that people possess intuitive moral sensibilities or moral foundations that apply when making moral judgments: Care/Harm, Fairness/Cheating, Loyalty/Betrayal, Respect/Authority, and Purity/Degradation. Haidt suggests that intuition, rather than reasoning, leads to moral judgment [3, 4]. The SIM raises four reasons to doubt the causality of reasoning in moral judgment: (a) There are two cognitive processes at work, reasoning, and intuition, being reasoning a process that has been overemphasized; (b) reasoning is usually motivated; (c) the reasoning process builds post hoc justifications, but we experience the illusion of objective reasoning; and (d) moral action covary with moral emotion more than with moral reasoning.

EthicApp is a collaborative application designed, implemented, and used to support ethical teaching [16–18] of instructional designs to support ethical discernment and moral reasoning of students under the rationalist-based moral psychology approach [5, 13]. It is available as a web application accessible from any device with a compatible web browser. EthicApp works in remote or face-to-face environments, synchronously or asynchronously. An ethical teaching activity supported by EthicApp can follow different instructional designs based on case-based learning and dilemma questions. Students can answer dilemma questions and semantic differentials, both individual and collaborative, associated with a case. In this research, we propose the use of EthicApp to support ethical teaching and instantiate instructional designs to support moral judgment based on **active learning** and **case-based learning**; this time, under the intuitionist psychology of the Moral Foundation Theory (MFT) [5, 14], and the Social Intuitionist Model (SIM) [3, 15]. It is essential to mention that the MFT has been applied very little as a theoretical framework for the design of ethical teaching activities [14, 15], and also in students who correspond to different countries that have different cultures, according to how the literature recommends apply [5, 19].

In this research, EthicApp is utilized as a scaffold for triggering moral deliberation as described by the SIM model. Two objectives drive this research. In the first place, we seek to analyze students' decisions in the context of a dilemma in academic ethics, considering their moral sensitivity as described by Moral Foundation Theory (MFT), and the Social Intuitionist Model (SIM). We seek to establish the existence of relationships between moral foundations and the stances that students take. For this, five samples of higher education students in four countries participate in the intervention. This follows recommendations in the literature regarding students from different cultures and social contexts [5]. Secondly, this study seeks to analyze the evolution of the decisions made by students throughout the successive activity phases, and instructional design operationalized through EthicApp, comprising different eliciting situations and anonymous small group discussions.

2 Theoretical Background

This section describes ethics teaching methodologies, and the advantages of active learning and case-based learning, both features incorporated in EthicApp. Likewise, the SIM and MFT used as a theoretical framework for the instructional design of the ethics activity supported by EthicApp are explained and applied in this study to students from four countries, each with its cultural characteristics.

2.1 Teaching Ethics – Active Learning with Case-Scenarios

According to [9], **active learning** in ethical teaching implies that the student actively participates in various pedagogical activities instead of having a passive role as in traditional methods where the teacher exposes concepts, recommendations, or examples, while students listen passively. Some examples of active learning type pedagogical activities are group presentations, role-play, case studies that include discussion and reflection, etc. The advantages of active learning are: (a) adding realism necessary to teach ethics; (b) facilitating more lasting and memorable teachings; and (c) increasing student motivation. In turn, according to Loeb (2015) [9], there are some possible disadvantages of active learning: (a) risks in giving control to students or not having the cooperation of students; (b) generating disgust in students due to the uncertainty associated with the expected results of the activity; (c) difficulty in its evaluation, mainly if it contains group parts; (d) exacerbate students' trust issues by making them reveal private information about themselves; and (f) possible difficulties with large groups of students.

In [19], three reasons are highlighted for using applications for mobile devices or apps that use Web 2.0 technology as scaffolding to support ethical teaching: (a) millennial and Z generations tend to use this type of technology more, and when used in applications to support ethical teaching, it makes their use attractive and satisfactory, which consequently favors support for critical thinking, problem-solving and the learning of values; (b) they help to democratize ethics teaching, since, in developing countries, they are used more frequently; (c) facilitate the introduction of ethical teaching tasks and activities through the support of simple applications; and (d) they allow the incorporation of more customized situations to the specific geographical and cultural context of the students, which in turn facilitates their comparison with other contexts.

According to [10], teaching methods based on case-based learning correspond to the following types: (a) case studies, or long narrative descriptions of real or hypothetical situations in which students are asked to identify, raise or resolve ethical dilemmas; (b) case-stories, which are stories that try to simulate the real world but are written by individuals in the classroom to show their points of view; and finally c) case-scenarios (vignettes), which are short stories made to express real-life situations concisely, thereby encouraging discussion and analysis of dilemmatic problems, for which there may be several possible solutions. **Case-scenarios** (a) combine unrealistic and real-life events to generate situations where there is no single correct answer, (b) deep subject knowledge is not necessary, (c) can be great brain teasers of discussion and cover complex and sensitive topics, (d) are readily applicable to people with different backgrounds who can quickly identify with the characters and participate in discussions from their perspective, and (e) there is enough detail to avoid unnecessary assumptions on the part of participants.

Regarding the identified disadvantages of active learning [9], EthicApp can mediate pedagogical tasks while supporting teachers' monitoring and control; it introduces anonymity among students when they collaborate, which reduces conflicting situations, decreases potential trust concerns, and lastly, it can be used with large cohorts synchronously. Regarding the benefits of Web 2.0 technologies indicated in [19], EthicApp is a web application that can be consumed from any device or operating system, thus meeting expectations of millennial and Z generations of students. Finally, EthicApp, can be used to instantiate case-based learning instructional designs [10].

2.2 Social Intuitionist Model

Haidt's Social Intuitionist Model (SIM) [3] includes actions of intuition, reasoning, judgment, as shown in Fig. 1, and are defined as follows: (a) **moral intuition**: the sudden appearance in consciousness of a moral judgment, planning process without any conscious knowledge of having gone through definite stages of decision making; (b) **moral judgment**, evaluations (good vs. bad) of the actions or character of a person that are made concerning a set of virtues; (c) **moral reasoning**, that is, conscious mental activity that transforms given information to arrive at a moral judgment, being a conscious process that is intentional, effortful and controllable.

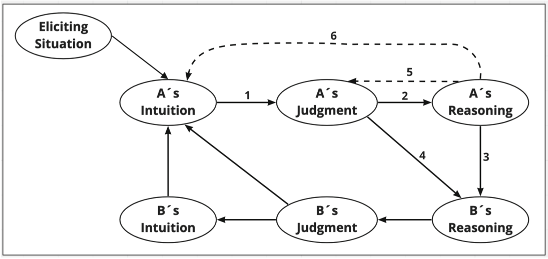


Fig. 1. Defined by [3]: 1 = the intuitive judgment link; 2 = the post-hoc reasoning link; 3 = the reasoned persuasion link; 4 = the social persuasion link; occurring less frequently are 5 and 6, the reasoned judgment link, and the private reflection link, respectively.

By the above, Haidt (2001) [3] incorporated into the SIM model the belief that morality stems from emotions that provide an immediate feeling of right or wrong. The model shows that moral judgment is a social process, not a rational cognitive individual act. It indicates that moral judgment results from intuitions or instantaneous moral foundations, which can be followed by moral reasoning; it focuses on moral intuition but does not eliminate the reasoning [3]. Links 1–4 reflect the process of making moral judgments and show the prominent role played by social intuition. The model demonstrates the belief that moral judgments are made through a process by which the person constantly assesses the environment and can quickly and intuitively make judgments; this is link 1 of the model. Link 2 shows that moral reasoning occurs after moral intuition and is an ad-hoc process that involves effort; that is, the person seeks support for their already made decision. Link 3, Reasoned Persuasion, demonstrates that people's reasoning (Link 2) can have a (causal) effect on others' intuitions. Haidt (2013) describes explicitly SIM as follows, [4]: "*In the social intuitionist view moral judgment is not just a single act that occurs in a single person's mind. It is an ongoing process, often spread out over time and over multiple people. Reasons and arguments can circulate and affect people…*" (p. 288).

Therefore, the moral judgments of friends, colleagues, and acquaintances can affect a person's decision. It is important to emphasize that the social intuition model does not rule out the reasoning process but instead acknowledges the complexity of making moral judgments based on intuition, reasoning, and social influences.

The model demonstrates a new aspect of moral decision-making, i.e., social intuition, and in doing so emphasizes the communications and interactions inherent in the process of addressing ethical issues. In this research, we aim to instantiate the SIM in the instructional design of the ethics teaching activity supported by EthicApp.

2.3 Moral Foundation Theory

In [5], there is consensus that morality comes from innate mental systems that result from an evolutionary process, the critical question being how many mental systems exist that explain morality. According to [5], Kohlberg's theory of the 70s, consisting of a single mental system oriented to whether something is fair or unfair, is shaped based on the perception of a person's social role, which is criticized for being insufficient. Given this, in the 1980s, critics of Kohlberg, such as Gilligan [20], introduced a new mental system in which the understanding that whether something is moral depends on how individuals understand, protect and respect others. Then, based on various problems presented by cultural variety and advances in the definition of mental systems, Haidt and Joseph posit that there are five moral foundations that explain people's moral decisions, giving rise to Moral Foundation Theory (MFT) [5]. MFT postulates that morality works innately through modules, which specialize in different situations that can arise in (social) life. **In addition, culture shapes morality**. Thus, a Moral Foundation can be understood as an unfinished morality, which must be developed and completed by the society in which it is inserted. Societies develop their morality based on one or more foundations. A subject's moral foundations described by MFT can be measured with the MFQ30 questionnaire, which comprises 30 items related to five factors describing people's morality (https://moralfoundations.org/questionnaires/).

According to Haidt [3], moral intuition can explain ethical decisions, giving four reasons: (1) both cognitive processes (intuition and reason) are involved in decision making, but only one of them is considered, (2) reasoning is influenced by emotion, (3) reasoning is a post-hoc construct to decisions we make, (4) there is greater covariance between moral intuition and moral action than between the latter and moral reasoning. Moral foundations are defined as follows [5, 21]:

- **Care/Harm (CRE):** Intuitions about avoiding emotional and physical damage to another individual.
- **Fairness/Cheating (FAI):** Intuitions about equal treatment and outcome for individuals and getting rewarded in proportion to their merit or contribution.
- **Loyal/Betrayal (LYL):** Intuitions about cooperating with in-groups and competing with outgroups.
- **Respect/Authority (RSP):** Intuitions about deference toward legitimate authorities and the defense of traditions are seen as providing stability and fending off chaos.
- **Purity/Degradation (PUR):** Intuitions about avoiding bodily and spiritual contamination and degradation

MFT has never been posited as an exhaustive list of moral foundations. The challenge to provide new foundations has been opened, such as Liberty/Oppression, Efficiency/Waste, or Ownership/Theft. Although there could be many foundations, MFT focuses on finding and defining the most important ones to understand interculturally, and behaviors based on morality [5].

To the present authors' best knowledge, research on how MFT could be used in ethics teaching in higher education has been scarce. Two investigations have been found in this regard. The first of these applied MFT to a group of accounting and business students, applying the MFQ30 questionnaire [14]. Among the conclusions reported by [14], it can be highlighted that after the analysis of invariances applied to the gender of the students, a greater correlation was observed in men in the contribution of the Loyalty/Betrayal; and a higher correlation in women with respect to Fairness/Cheating and Loyalty/Betrayal. Likewise, greater sensitivity was identified in the moral foundations of Fairness/Cheating and Loyal/Betrayal in accounting students than in business students. Finally, they suggest considering the difference between the moral reasoning model and moral intuition. The second investigation [15], presented Haidt's SIM [2] as a new approach to understand how moral judgments are made, that is, quickly and intuitively. These authors explored the applicability of SIM using an ethical case in five different accounting courses, which provided them with a means to improve accounting ethics education by allowing students to exercise their intuition as an initial stage of the decision-making process.

2.4 EthicApp

Based on the analysis of the literature on ethics education, we developed an application called EthicApp to support teaching of ethics [16–18]. EthicApp includes functions to author and run instructional designs for students' active learning, with case-based scenarios. Its design principles are: (a) easy integration in any course; the teacher does not need to make significant changes in their study plan or pedagogical methodology to use the application, allowing them to reuse their materials and content; (b) easy to use, since it has a simple and minimalist interface, with the purpose of minimizing teacher training efforts and facilitating quick and intuitive adoption; (c) promotes discussion, reflection, and discernment by capturing quantifiable assessments, judgments, and written ethical arguments from students, when analyzing cases and ethical dilemmas; (d) incorporates anonymity during collaborative interaction among students so that they can express their opinions and make authentic ethical judgments while reducing conflict anxiety with their peers; (e) supports discussion and argumentation through real-time collaborative chat discussions; (f) domain independence, since EthicApp can be used in any professional area in which cases or ethical dilemmas can be analyzed; (g) offers efficient management of the information generated by the students, presenting it to the teacher through a simple and easy-to-understand dashboard, thus facilitating monitoring the pedagogical activity; (h) allows combining individual work and collaborative work, which can be carried out based on successive phases prompting for individual assessment of the ethical case, re-elaboration of judgments, and synchronous collaborative discussion among students to explain justifications, argue and debate; (i) incorporates flexibility in its configuration, in order to support diverse instructional designs; (j) supports different types of devices and form factors (i.e., smartphones, tablets, laptops, etc.).

3 Method

3.1 Samples

Six samples of engineering and business students from higher education institutions in Chile (CL), Colombia (CO), Ecuador (EC), and Mexico (MX) participated in this study (see Table 1).

Table 1. Participating institutions, courses, sample sizes, student participation in EthicApp activities 1–3, and responses to MFQ30.

Country	Institution	Courses	N	Participants per activity			MFQ30
				Act. 1	Act. 2	Act. 3	
Chile (CL)	Universidad de Chile (UCHILE)	Organizational Design and Planning	102	102 (100%)	98 (96%)	55 (54%)	59 (58%)
	Universidad de los Andes (UANDES)	Professional Ethics Seminar for Engineering	56	52 (93%)	56 (100%)	32 (57%)	36 (64%)
Colombia (CO)	Universidad del Cauca (UNICAUCA)	Computer-Supported Collaborative Learning	20	20 (100%)	19 (95%)	19 (95%)	17 (85%)
Ecuador (EC)	Universidad Técnica Particular de Loja (UTPL)	Ethics and Morality	36	31 (86%)	36 (100%)	21 (58%)	25 (69%)
Mexico (MX)	Benemérita Universidad Autónoma de Puebla (BUAP)	Software Quality	19	32 (89%)	29 (81%)	30 (83%)	27 (75%)
		Agile Project Management	17				
		Total	249	237	238	157	164

Only in UANDES (CL) and in UTPL (EC) the intervention with EthicApp was carried out in a dedicated ethics course. In the other institutions, the case was carried out in disciplinary courses of the curricula.

3.2 Instructional Design and Ethical Case

Instructional Design. The instructional design includes the realization of three activities with EthicApp (see Fig. 2). The first activity consists of intuitively responding to a vignette that is related to the analyzed ethical case, however, the students have not yet read the case. The second activity is carried out synchronously, online or in face-to-face, and consists of individual (see a21, a22, a23, and a25 in Fig. 2) and collaborative work

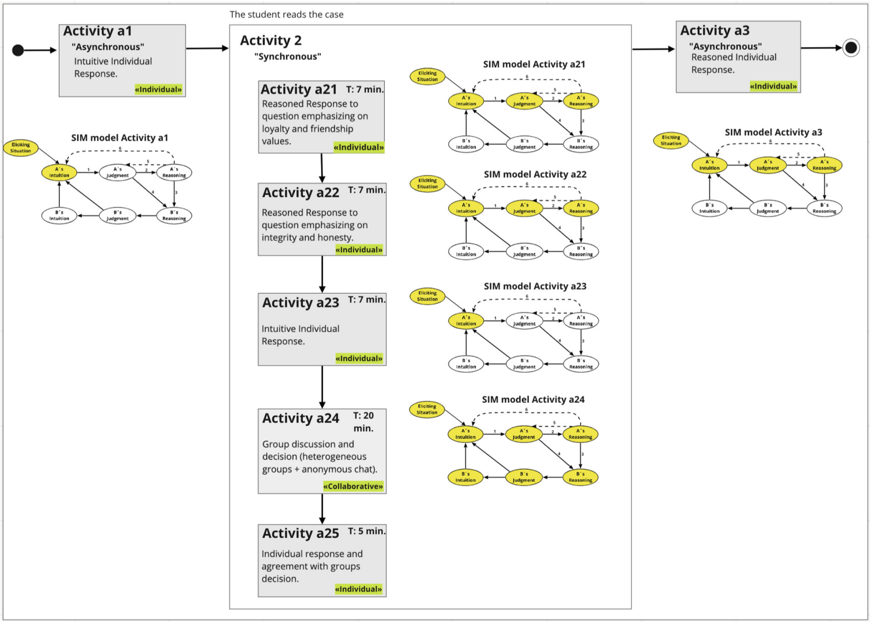


Fig. 2. Instructional design comprising asynchronous activity 1 (pre-test), synchronous activity 2 (intervention), and asynchronous activity 3 (post-test).

(see a24 in Fig. 2), in which specific phases of the SIM model (see Fig. 1) are designed to be activated, based on the kind of question, the instructions given to the students, and the interaction mode. The third activity (see a3 in Fig. 2) is asynchronous, two to three weeks after the second activity.

Ethical Case. The ethical case in this activity was created by the present authors. It involves Sebastián, a Business student who struggles with his Accounting course. Sebastian needs to pass his final exam to stay in college, since he depends on a scholarship and will lose the chance to continue studying if he fails the course. Throughout the story, written in the first person, the reader assumes the role of Sebastian's best friend, who lived with him and his family during the COVID-19 pandemic, and accompanies him in the situation he is experiencing.

The text of the case emphasizes a series of values that the reader will be able to rank according to their moral sensitivity, such as individual responsibility and merit, honesty and intellectual integrity, and loyalty and friendship. Individual responsibility and merit are expressed in the case text through the reality of Sebastian's family; a middle-class family that has managed to progress thanks to the effort and merit of their parents. Sebastian's father advises Sebastian to "work hard and see the results". Intellectual honesty and integrity are treated through a reflection that the narrator makes about the negative of academic plagiarism. Finally, loyalty and friendship are emphasized by the close relationship Sebastián and he have with his friend – the reader – and the loyalty they have had for each other for a long time.

Table 2. Case questions and response types per activity and phase.

Act.	Indications and question text	Response*
a1	Sebastián, a very good friend of yours, needs to pass the accounting course in order to stay in college; he is struggling with this and asks you to help him take the online exam. What would you do?	1–7 SDS
a21	[Provide arguments to justify your answer]. Sebastian needs to pass the exam and I owe him a lot for his loyalty and friendship. Even he and his family took me in for a few months during the pandemic. Also, if he does not finish the course well, he will have serious financial problems to continue studying. What would you do?	1–6 SDS
a22	[Provide arguments to justify your answer]. With Sebastián, we have been very critical of those who cheat to advance in their courses. We always fight their hypocrisy. We try very hard to pass courses and we really want to be good professionals. If we want to be good professionals and be up to the task, what I must do is:	
a23	[Provide arguments to justify your answer]. Given Sebastian's situation, my decision would be:	
a24	[Agree on a consensual response, by discussing anonymously via chat aspects that you have already considered or others that you have not considered in the previous questions. That is, everyone must choose the same numerical value in the answer, even though each one can provide different supporting arguments] Given Sebastian's situation, my decision would be:	
a25	[Please provide arguments to justify your answer.] How much do you agree with your group's answer in the previous stage?	
a3	Sebastián, a very good friend of yours, needs to pass the accounting class to be able to stay in college. He is very complicated, and he asks you to help him take the online exam. What would you do?	1–7 SDS

* All questions require a written response in addition to the semantic differential scale. Act = Activity

Table 2 shows case questions in relation to the instructional design. Questions have a Semantic Differential Scale (SDS) component with values from 1 to 7 (to allow expressing indifference, neutrality or insecurity in the intermediate value 4), or values from 1 to 6 (to force a response towards one of the two poles of the semantic differential scale). In either case, the poles of the SDSs are always 'Help Sebastian with his exam', and 'Not help him with his exam'. The questions also ask the student to write a justification for their response to the SDS.

3.3 Procedure

Each institution carried out the activity with EthicApp based on the case presented above, in a period of four weeks. In the first week, invitations were sent via email and/or as announcements in the Learning Management System (LMS) to students to sign up on

EthicApp and enter the first activity, along with invitations to an informed consent form, and an adaptation to Spanish of the MFQ30 questionnaire (https://moralfoundations.org/questionnaires/). Responses to MFQ30 were collected using Google Forms. After the first activity, the Sebastian case text was distributed to the students. In activity 2, the students synchronously online or face-to-face, carried out the five phases, in a single session and with the presence of their respective teacher. Two to three weeks after activity a2, students had to answer a final question in activity a3.

3.4 Analyses

The data from the MFQ30 questionnaire and the responses captured with EthicApp were analyzed in an environment based on R version 4.2, with packages dplyr, ggplot2, psych, and lavaan, among others. A descriptive analysis of the MFQ data was first carried out, and then an exploratory analysis looking for correlations between the components of MFQ30, and students' answers to the question in activity a1, in which answers was expected to be intuitive. Next, students' responses were analyzed throughout the successive activities of the instructional design, observing frequencies of students who maintained or changed stance. Modifications of stances were classified according to their characteristics, e.g., if the student moved their response closer to or further from one of the poles, or if the student's response swung from one pole to another. Finally, we constructed a path model to examine to what extent the response in one activity predicts the response in successive activities.

4 Results

4.1 Student Participation

Table 1 summarizes students' participation in EthicApp activities, and in responses to the MFQ30 questionnaire. In activities a1 and a2, participation was consistently above 80%. Since activity a3 was carried out by the students two to three weeks after activity a2, and asynchronously, participation rate was lower; between 50% and 60% in institutions in Chile and Ecuador, and over 80% in Mexican and Colombian institutions. MFQ30 response rates were also variable, with a minimum of 58% in UCHILE, and a maximum of 85% in UNICAUCA.

4.2 Moral Foundations

Figure 3 shows distributions of MFQ30 components in each subsample by country. It is observed that the FAI, CRE and LYL components have a small variance compared to RSP and PUR. By country, Chile has the highest variance in LYL, PUR, RSP variables and in the sum of MFQ30 scores (see Fig. 4). The other subsamples have lower variances, attributable to smaller sizes.

Table 3 shows descriptive statistics for MFQ30 by component, along with Cronbach's alpha. The components of the Spanish version of MFQ30 used in this study do not reach the acceptable internal consistency of 0.7, except for RSP and PUR factors. Thus, for

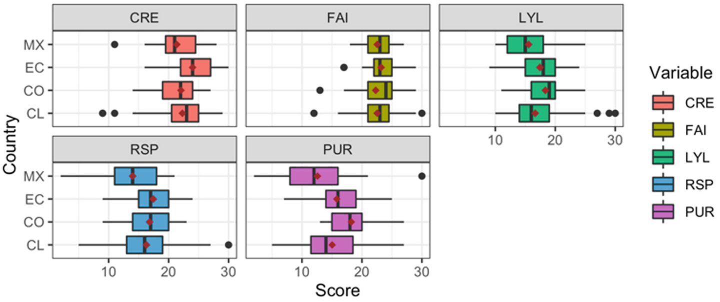


Fig. 3. Distribution of MFQ30 scores per factor and country subsample.

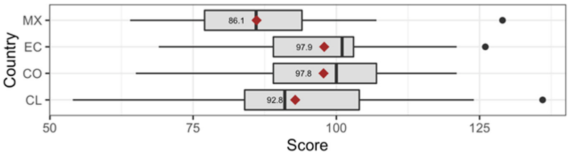


Fig. 4. Distribution of MFQ30 total scores per country subsample.

future studies, alternative adaptations of the instrument to Spanish should be proposed and validated.

Figure 4 shows the MFQ30 total score distributions by country, indicating the mean of each subsample. The samples from Colombia and Ecuador are very similar, while Chile has a higher variance, and Mexico the lowest mean and lowest variance.

Table 3. MFQ30 descriptive statistics.

Factor	Min	Median	Max	*M*	*SD*	Cronbach's α
CRE (Care/Harm)	9.0	23.0	30.0	22.4	4.0	.59
FAI (Fairness/Cheating)	12.0	23.0	30.0	22.6	3.1	.57
LYL (Loyalty/Betrayal)	9.0	17.0	30.0	16.8	4.1	.46
RSP (Respect/Authority)	2.0	16.0	30.0	16.1	5.0	.70
PUR (Purity/Degradation)	2.0	15.0	30.0	15.1	5.4	.69

4.3 Moral Foundations vs. Students' Judgment

Figure 5 (left) presents a correlogram based on Spearman's correlation, which includes MFQ30 factors and students' response scores in activity a1. The complete data of students who participated in a1 and gave a valid and complete response to the MFQ30 questionnaire includes 130 cases. No significant correlations are observed at the 0.05 level between MFT factors and the students' response in activity a1.

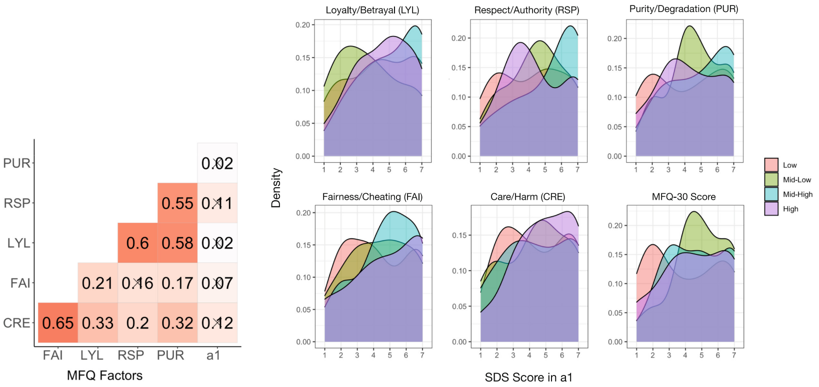


Fig. 5. Left: Correlogram including MFQ30 factor scores and students' response scores in a1. Non crossed out correlations are significant at the 0.05 level. Right: Densities of MFT components and MFQ30 total score vs response in a1.

In order to visually explore relationships between MFQ30 components and student responses in activity a1, Fig. 5 (right) shows density plots of MFQ30 components versus SDS values in a1. The density curves correspond to four levels of the variables (i.e., Low, Mid-Low, Mid-High, and High) according to quartiles in their distribution (i.e., Q1, Q2, Q3, and Q4, respectively). We observe that in the FAI and CRE variables there is a difference between Low and High level students in relation to the score they give on the SDS. The students in the High group are mostly inclined not to help Sebastián, while in the Low group there is more willingness to help him.

Figure 6 shows proportions of students who were inclined to help, not help and who chose the midpoint on the SDS, in activities a1 and a3, segmented by MFQ30 score (left), levels of CRE (center), and levels of FAI (right). A higher proportion of High level students in CRE were inclined not to help Sebastián, while in the Low level the proportion is somewhat lower. In activity a3, the proportion of students choosing to help Sebastián increased at all levels of the CRE variable, and very noticeably at the High level. Measurements of the FAI variable result in similar conclusions. Considering the total MFQ30 score (see Fig. 6-left), the Mid-Low group shows the greatest change in stance between activities a1 and a3. In a1, the decision not to help was dominant in this group, while in a3 it decreased considerably, due to a proportional increase in students opting for both help and the mid-point option.

4.4 Evolution of Students' Responses

Figure 7 shows relative frequencies of students' change of stance throughout the phases of the instructional design. Between a1 and a21 (Fig. 7a), there were about 27% of students who swung from 'HELP' to 'NOT HELP' and vice-versa. Seventeen percent of the complete sample had answered midpoint in a1 and switched to 'HELP' or 'NOT HELP' in a21. Notably, in the subsample from Ecuador, a third of the students transitioned from 'NOT HELP' to 'HELP', and in the Mexican subsample, 40% of the students maintained their original stance unchanged.

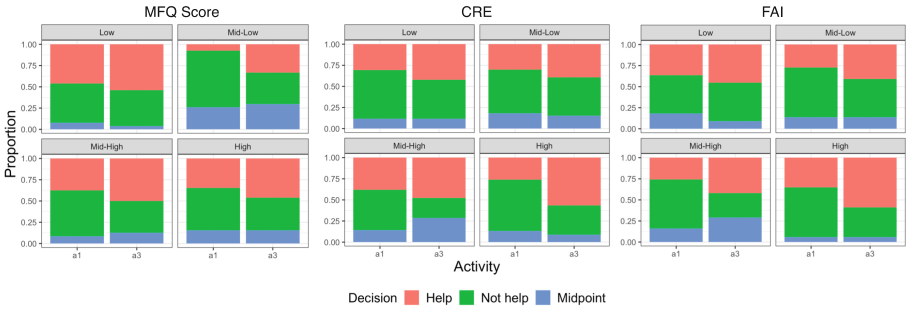
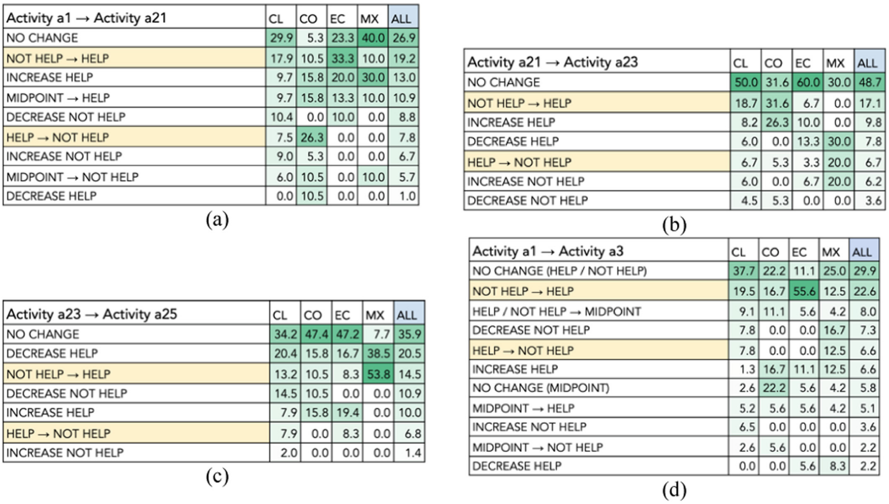


Fig. 6. MFQ30 total score (left), CRE (center), and FAI (right) levels vs. decisions in activities a1 and a3.



Activity a1 → Activity a21	CL	CO	EC	MX	ALL
NO CHANGE	29.9	5.3	23.3	40.0	26.9
NOT HELP → HELP	17.9	10.5	33.3	10.0	19.2
INCREASE HELP	9.7	15.8	20.0	30.0	13.0
MIDPOINT → HELP	9.7	15.8	13.3	10.0	10.9
DECREASE NOT HELP	10.4	0.0	10.0	0.0	8.8
HELP → NOT HELP	7.5	26.3	0.0	0.0	7.8
INCREASE NOT HELP	9.0	5.3	0.0	0.0	6.7
MIDPOINT → NOT HELP	6.0	10.5	0.0	10.0	5.7
DECREASE HELP	0.0	10.5	0.0	0.0	1.0

(a)

Activity a21 → Activity a23	CL	CO	EC	MX	ALL
NO CHANGE	50.0	31.6	60.0	30.0	48.7
NOT HELP → HELP	18.7	31.6	6.7	0.0	17.1
INCREASE HELP	8.2	26.3	10.0	0.0	9.8
DECREASE HELP	6.0	0.0	13.3	30.0	7.8
HELP → NOT HELP	6.7	5.3	3.3	20.0	6.7
INCREASE NOT HELP	6.0	0.0	6.7	20.0	6.2
DECREASE NOT HELP	4.5	5.3	0.0	0.0	3.6

(b)

Activity a23 → Activity a25	CL	CO	EC	MX	ALL
NO CHANGE	34.2	47.4	47.2	7.7	35.9
DECREASE HELP	20.4	15.8	16.7	38.5	20.5
NOT HELP → HELP	13.2	10.5	8.3	53.8	14.5
DECREASE NOT HELP	14.5	10.5	0.0	0.0	10.9
INCREASE HELP	7.9	15.8	19.4	0.0	10.0
HELP → NOT HELP	7.9	0.0	8.3	0.0	6.8
INCREASE NOT HELP	2.0	0.0	0.0	0.0	1.4

(c)

Activity a1 → Activity a3	CL	CO	EC	MX	ALL
NO CHANGE (HELP / NOT HELP)	37.7	22.2	11.1	25.0	29.9
NOT HELP → HELP	19.5	16.7	55.6	12.5	22.6
HELP / NOT HELP → MIDPOINT	9.1	11.1	5.6	4.2	8.0
DECREASE NOT HELP	7.8	0.0	0.0	16.7	7.3
HELP → NOT HELP	7.8	0.0	0.0	12.5	6.6
INCREASE HELP	1.3	16.7	11.1	12.5	6.6
NO CHANGE (MIDPOINT)	2.6	22.2	5.6	4.2	5.8
MIDPOINT → HELP	5.2	5.6	5.6	4.2	5.1
INCREASE NOT HELP	6.5	0.0	0.0	0.0	3.6
MIDPOINT → NOT HELP	2.6	5.6	0.0	0.0	2.2
DECREASE HELP	0.0	0.0	5.6	8.3	2.2

(d)

Fig. 7. Relative frequencies (percentages) of transition types in students' decisions (a) from activity a1 to a21, (b) from activity a21 to a23, (c) from activity a23 to activity a25, and from activity a1 to a3, broken down by country subsamples and total sample (ALL).

In activity 2 (Fig. 7b), about half (i.e., 48.7%) of the students maintained their response on the SDS when comparing a21 and a23. This trend is higher in Chilean (50%) and Ecuadorian (60%) subsamples. Seventeen percent of the complete sample changed their position from 'NOT HELP' to 'HELP', and only 6.7% changed in the opposite decision. About fifty percent of the Mexican sample swung from 'HELP' to 'NOT HELP' (20%), or decreased willingness to help (30%).

A comparison between students' responses in a23 and a25 (Fig. 7c), that is, considering students' individual response in a25 after the group discussion in a24, in which students mostly moderate their postures, as they are compelled to come to an agreement, it is observed that just over a third of students of the complete sample (35.9%) kept their position unchanged. However, 14.5% swung from 'NOT HELP' to 'HELP', and 6.8% in the opposite way. The rest changed the intensity of the stance adopted in a23, mostly in a decreasing fashion (i.e. ‘DECREASE HELP’ with 20.5%, and ‘DECREASE NOT HELP’ with 10.9%). In the Mexican subsample, 53.8% of the students swung from ‘NOT HELP’ to ‘HELP’ in this transition.

Figure 7d shows that between a1 and a3, 29.9% of the complete sample maintained their posture unchanged, and another 29.2% of the students swung from ‘NOT HELP’ to ‘HELP’ and vice-versa (22.6% and 6.6%, respectively). In the Ecuadorian sample, 55.6% of students’ decisions evolved this way. Eight percent of the complete sample changed their stance from ‘NOT HELP’, or ‘HELP’, to ‘MIDPOINT’. It is apparent that throughout the activity, students tend to increase their willingness to help Sebastian, which is consistent with results in Fig. 6.

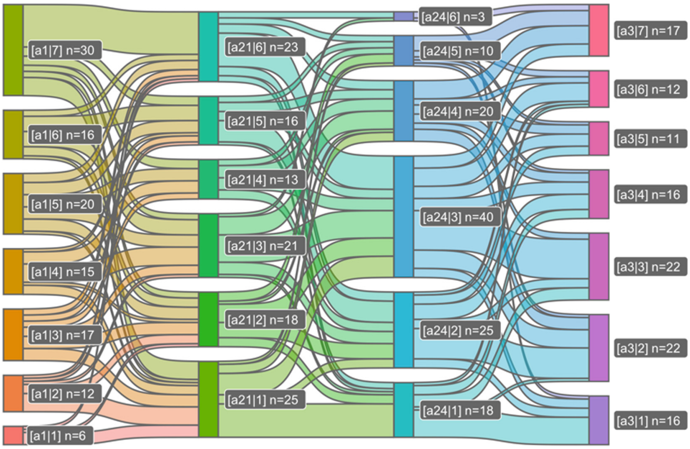


Fig. 8. Sankey diagram depicting evolution of students’ SDS scores throughout activities a1, a21, a24 and a3, comprising 116 complete cases.

Figure 8 shows a Sankey diagram for 116 complete cases, created with the ggsankey R package (https://github.com/davidsjoberg/ggsankey). In the diagram it is possible to observe how students’ decisions vary from one phase to the next, considering that the first column shows the values 1–7 of the SDS in a1, the two middle columns show values 1–6 from the SDSs in activities a21 and a24, and the last column shows values 1–7 from a3. Students’ decisions do change between successive phases, which confirms the occurrence of results that align with the rationale of the instructional design. Evolution of students’ decisions shows that there is influence from the antecedents in the text of the case (i.e., contrast between a1 and a21), the social influence and the need to reach a group consensus (a24), and the reflection that the students carry out in a3, weeks after the activity a24 was carried out.

Finally, Fig. 9 shows a path model created with lavaan [22] and lavaanPlot R packages, with significant standardized coefficients at levels 0.01 (**) and 0.001 (***). The path model shows how response values to the SDSs in an activity predict responses in successive activities. For example, the decision the student makes in a1 has limited predictive power relative with regard to the decision in a3 (i.e., non-significant std. Path coefficient of 0.12). A Wilcoxon Signed-Rank Test indicated that the median post-test (a3) ranks were statistically significantly higher than the median pre-test (a1) ranks V = 2973.5, $p < 0.001$. Contrastingly, the process that the student experiences in activity 2, mediated by the reading of the case (a21), reflective questions (a22 and a23), group discussion (a24), and individual response (a25) influences considerably (i.e., highly significant std path coefficient of .67) on what the student finally answers in a3.

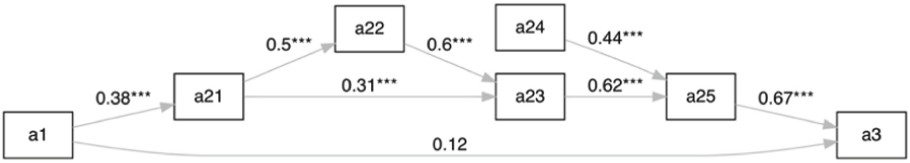


Fig. 9. Path model based on students' responses to SDS in activities a1, a2 and a3.

5 Discussion

The experience analyzed in this study provides evidence that the instructional design scaffolded by EthicApp elicits heterogeneous positions of students when facing an ethical dilemma at the outset of the process. However, students' postures can vary as a result of different cognitive and social stimuli that arise in the successive stages of instructional design, even to the point where a student can change their stance to the opposite side. This was observed in 28.2% of the cases observed with full participation in activities a1 and a3 of the instructional design (N = 137).

Although it has not been possible to predict students' responses to the initial question of the case based on MFT components measured with MFQ30, students appear to define an initial position guided by their intuitions and biographical experiences, and such initial position evidences a certain way preferred by the student to prioritize values. Some students place friendship and loyalty with Sebastian at the top of the hierarchy, while other students are inclined to prioritize academic integrity and acting in good faith over the pedagogical and formative character on which the evaluation processes are based.

A central objective of the intervention with EthicApp is to make students aware of the existence of different positions on the part of their classmates. This is in line with current normative conceptions about the orientation that an ethical education should have, particularly in the field of educational engineering and business [1, 6]. However, from the pedagogical point of view, reflection on the consequences and ethical implications of alternative value hierarchizations is essential. Therefore, the teacher's role, informed by EthicApp, should consist of guiding students on the most correct solution to the ethical dilemma and the rationale for the solution, contrasting this solution with the positions and justifications expressed by the students in the successive phases of the instructional design.

It is not possible to assert that the above can be fully resolved through the proposed instructional design, and arguably, it is necessary to propose and evaluate other complementary interventions that allow students to continue their reflection and internalize ethical reasoning schemes in their cognitive structures, that are functional to the situation of ethical deliberation in similar circumstances in their lives as students, and in their professional future.

6 Conclusions and Future Work

In this paper we have presented the design of a scaffolding for ethics teaching based on EthicApp and the Social Intuitionist Model, and a case about academic ethics that can be implemented in both disciplinary and dedicated courses in ethics, in business and STEM curricula. We conducted an exploratory study with measurements based on MFT, and involving samples of students in five higher education institutions in four countries, with up to 238 students participating in the activities comprised in the intervention.

While measurement of MFT through MFQ30 did not appear to correlate nor explain students' intuitive response to eliciting situations posed by the instructional design, the samples studied from four different countries exhibit minor differences in Care/Harm, Fairness/Cheating, and Loyalty/Betrayal factors, while greater differences and variances are observed with regard to Respect/Authority and Purity/Degradation factors. In the first activity of the intervention, a greater proportion of students with high scores in Harm/Care and Fairness/Cheating factors were reluctant to help a struggling classmate and friend sit an online exam than students with lower scores in these variables. However, the process in the second activity of the instructional design, mediated by reasoned responses informed by factual knowledge (i.e., in the actual case text), and social interaction, significantly influenced students' responses in the third phase of the instructional design. Furthermore, a significant difference was observed between students' responses to questions in the first and last activities of the intervention. This indicates that students' ethical reasoning, initially guided by intuition as [3, 4] prescribes, can be indeed be influenced by facts, reflective questions, and social interaction.

As for future work, an important question that stems from this research is how students' psychological processes linked to ethical schema building could be scaffolded, and oriented, both pedagogically – taking the teacher's role into account – and technology-wise, towards identifying and standing for solutions to ethical dilemmas that are based on virtue and bring greater good.

References

1. Pfeffer, J., Fong, C.T.: The business school 'business': some lessons from the US experience. J. Manage. Stud. **41**(8), 1501–1520 (2004)
2. Holsapple, M.A., et al.: Framing faculty and student discrepancies in engineering ethics education delivery. J. Eng. Educ. **101**(2), 169–186 (2012)
3. Haidt, J.: The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychol. Rev. **108**(4), 814 (2001)
4. Haidt, J.: Moral psychology for the twenty-first century. J. Moral Educ. **42**(3), 281–297 (2013)
5. Graham, J., et al.: Moral foundations theory: the pragmatic validity of moral pluralism. In: Advances in Experimental Social Psychology, pp. 55–130. Elsevier (2013)
6. Hess, J.L., Fore, G.: A systematic literature review of US engineering ethics interventions. Sci. Eng. Ethics **24**(2), 551–583 (2018). https://doi.org/10.1007/s11948-017-9910-6

7. Poje, T., Zaman Groff, M.: Mapping ethics education in accounting research: a bibliometric analysis. J. Bus. Ethics **179**, 451–472 (2022). https://doi.org/10.1007/s10551-021-04846-9
8. Cornelius, N., Wallace, J., Tassabehji, R.: An analysis of corporate social responsibility, corporate identity and ethics teaching in business schools. J. Bus. Ethics **76**(1), 117–135 (2007). https://doi.org/10.1007/s10551-006-9271-6
9. Loeb, S.E.: Active learning: an advantageous yet challenging approach to accounting ethics instruction. J. Bus. Ethics **127**(1), 221–230 (2015). https://doi.org/10.1007/s10551-013-2027-1
10. Jeffries, C., Maeder, D.W.: Using instructional and assessment vignettes to promote recall, recognition, and transfer in educational psychology courses. Teach. Educ. Psychol. **1**(2), n2 (2006)
11. Impagliazzo, J., Gorgone, J.: Professional accreditation of information systems programs. Commun. Assoc. Inf. Syst. **9**(1), 3 (2002)
12. Kulturel-Konak, S., et al.: Assessing professional skills in STEM disciplines. In: 2013 IEEE Integrated STEM Education Conference (ISEC). IEEE (2013)
13. Kohlberg, L.: Stage and sequence: the cognitive-developmental approach to socialization. In: Handbook of Socialization Theory and Research, vol. 347, p. 480 (1969)
14. Andersen, M.L., Zuber, J.M., Hill, B.D.: Moral foundations theory: An exploratory study with accounting and other business students. J. Bus. Ethics **132**(3), 525–538 (2015). https://doi.org/10.1007/s10551-014-2362-x
15. Andersen, M.L., Klamm, B.K.: Haidt's social intuitionist model: what are the implications for accounting ethics education? J. Account. Educ. **44**, 35–46 (2018)
16. Alvarez, C., Zurita, G., Baloian, N., Jerez, O., Peñafiel, S.: A CSCL script for supporting moral reasoning in the ethics classroom. In: Nakanishi, H., Egi, H., Chounta, I.-A., Takada, H., Ichimura, S., Hoppe, U. (eds.) CRIWG+CollabTech 2019. LNCS, vol. 11677, pp. 62–79. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-28011-6_5
17. Alvarez, C., Zurita, G., Carvallo, A., Ramírez, P., Bravo, E., Baloian, N.: Automatic content analysis of student moral discourse in a collaborative learning activity. In: Hernández-Leo, D., Hishiyama, R., Zurita, G., Weyers, B., Nolte, A., Ogata, H. (eds.) CollabTech 2021. LNCS, vol. 12856, pp. 3–19. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-85071-5_1
18. Alvarez, C., et al.: A social platform for fostering ethical education through role-playing. In: Factoring Ethics in Technology, Policy Making, Regulation and AI, p. 107 (2021)
19. Montiel, I., et al.: New ways of teaching: using technology and mobile apps to educate on societal grand challenges. J. Bus. Ethics **161**(2), 243–251 (2020). https://doi.org/10.1007/s10551-019-04184-x
20. Gilligan, C.: In a Different Voice: Psychological Theory and Women's Development. Harvard University Press (1993)
21. Haidt, J., Joseph, C.: The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. The Innate Mind **3**, 367–391 (2007)
22. Rosseel, Y.: lavaan: an R package for structural equation modeling. J. Stat. Softw. **48**, 1–36 (2012)

Development of Toys for Determining Behavioral Imitation During Parent-Child Interactions

Takashi Numata[1](✉), Masashi Kiguchi[1], Hisanori Matsumoto[1], Atsushi Maki[1], Masako Kawasaki[2], and Toshiko Kamo[2]

[1] Hitachi, Ltd., Higashi-Koigakubo 1-280, Kokubunji, Tokyo, Japan
takashi.numata.rf@hitachi.com
[2] Japan PCIT Training Center, Wakamatsu-Cho 9-4, Shinjuku, Tokyo, Japan

Abstract. The number of parents who are troubled by their children's mental and behavioral problems has been increasing. Parent-Child Interaction Therapy (PCIT), which is a psychotherapy based on play therapy, has been used to help such parents and children by improving the quality of interactions between them through toy play. In PCIT, the number of behavioral imitations during toy play is one of the key indicators to evaluate parenting skills. However, it is difficult to manually evaluate behavioral imitations because it is necessary to simultaneously observe the parent and child manipulating toys to determine whether and to what extent the parent and child imitate each other's behaviors. We developed two types of toys; toy blocks equipped with sensors and digital drawing application, which enable us to visualize toy manipulations of parents and children and determine behavioral imitations between them. Toy manipulations during parent-child interactions were visualized, and the number and/or the extent of behavioral imitations were determined from the toy-manipulation data, which were obtained using our toys. The results indicate that our toys should have the potential to determine the number and extent of behavioral imitations during parent-child interactions. Thus, these toys should be useful to evaluate parenting skills.

Keywords: Digital toy · Parent-child interaction · Behavioral imitation · Correlation analysis · Dynamic time warping

1 Introduction

1.1 Background

Child maltreatment has been a serious global problem, and the COVID-19 pandemic and physical distancing have greatly increased the risk of intra-family violence and child abuse [1]. Therefore, the number of parents having to deal with the mental and behavioral problems of their children is increasing. Parent-Child Interaction Therapy (PCIT) is an evidence-based psychotherapy to support such parents [2]. In PCIT, parents wear earphones and are coached directly by a therapist in the observation room, while playing with their child in the playroom. PCIT helps parents and children improve the quality of their interactions and solve mental and behavioral problems, and improvements in

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 279–286, 2022.
https://doi.org/10.1007/978-3-031-20218-6_19

parent-child relationships has been demonstrated [3, 4]. An Internet-delivered version of PCIT has also been developed to overcome physical distancing [5].

The PCIT therapist observes the behavior of parents and children during parent-child interactions through toy play and guides them to improve their relationships. In PCIT, one of the key indicators of parenting skills and trust relationship between parents and children is the frequency of behavioral imitation between them. Behavioral imitation is useful for inducing positive emotions and building empathy toward the imitated person [6–8], thus helps improve parent-child relationships. However, quantitative evaluation of behavioral imitation is particularly difficult because it is necessary to simultaneously observe parent-child toy manipulation and evaluate whether behavioral imitation occurs and to what extent. Therefore, it is useful to visualize and quantitatively evaluate the behavioral imitation between parents and children during toy play.

In previous studies, various toys, such as stuffed toys and building blocks equipped with sensors, have been developed [9–11]. These toys have advantages in measuring toy manipulations. However, they have not been used by multiple users and developed to visualize the interactions of toy manipulation between parents and children, such as behavioral imitation. Therefore, it is useful to develop such toys that enable us to visualize and quantitatively evaluate behavioral imitation during parent-child interactions.

1.2 Objective

The objective of this study was to develop toys that enable us to visualize and quantitatively evaluate behavioral imitation between parents and children during parent-child interactions. This is the first study to develop toys for visualization and quantitatively evaluation of behavioral imitation during PCIT. Specifically, we focused on block playing and drawing, which are the two main types of play in PCIT, to measure toy manipulation. We then evaluated the possibility of visualization and quantitative evaluation of behavioral imitation during toy play by using these toys.

2 Development and Evaluation of Sensor-Equipped Blocks

2.1 Development of Digital Blocks

In PCIT, it is necessary to use toys that can be operated naturally by parents and children. Therefore, we developed blocks equipped with sensors that can be used in the same manner as conventional toy blocks.

A 25 × 25 × 10 mm three-axis accelerometer (2525A, Mono wireless) was installed inside each 60 × 60 × 80 mm block (Dekoboko Block, Yuai-Gangu) (Fig. 1(A)). A sensor was installed inside the cavity of each block (Fig. 1(B)). The acceleration sensors can transmit acceleration measurement data via Zigbee. We developed a system to visualize the multiple acceleration sensors' data by receiving the data on a laptop via a USB dongle (MONOSTICK-B, Mono wireless) (Fig. 1(C)). This system enables the participants to play with the blocks in the same manner as conventional blocks and visualize the movements of the blocks.

2.2 Experimental Procedure

In the experimental environment, a pair of healthy male participants were asked to sit side-by-side, and eight blocks were placed in front of each participant (Fig. 1(D)). To evaluate the possibility of visualizing behavioral imitation by using these blocks, the pair was asked to perform an imitation task by using the blocks. Data from the participants were obtained following receipt of written informed consent.

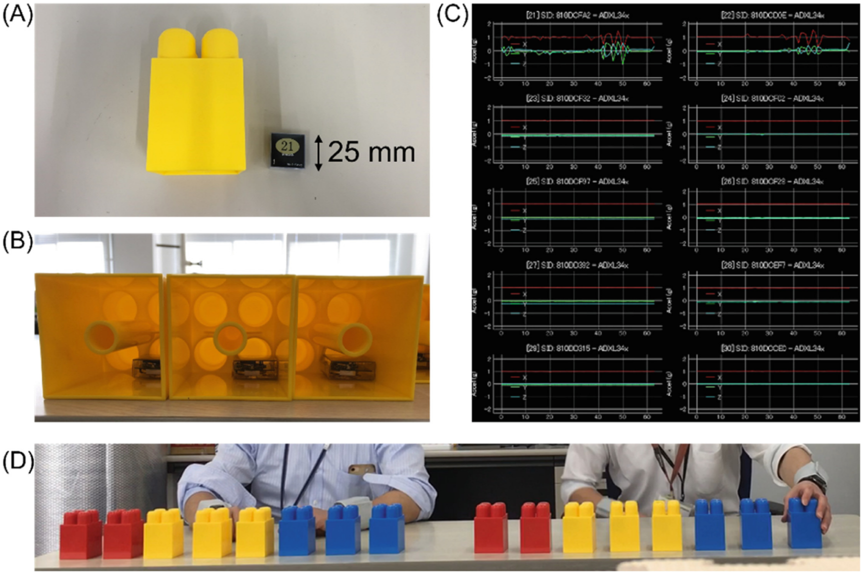


Fig. 1. Developed blocks with sensors. (A) Sizes of block and accelerometer used in this study. (B) Installed accelerometers in blocks. (C) Motion visualization system of blocks by using the multiple 3-axis acceleration sensors' data installed inside the blocks. (D) Initial positions of participants and initial placement of blocks during experiment.

The imitation task involved a manipulator and imitator. The role of the manipulator was to manipulate the digital blocks in an arbitrary order one at a time to assemble an arbitrary block structure. The role of the imitator was to imitate that block manipulation of the manipulator, assembling the same block structure in the same order as the manipulator's. The blocks were numbered from one to eight (Block A1-A8 for the manipulator and Block B1-B8 for the imitator).

To detect who manipulated the blocks (the manipulator or imitator) and which hand manipulated the blocks (left or right) to evaluate behavioral imitation, two types of accelerometers were attached to the wrists of both the manipulator and imitator. One is a highly sampled three-axis accelerometer (TSND151, ATR-Promotions) with a sampling rate of 1,000 Hz. They were used to detect who manipulated the blocks and with which hands. The other accelerometer was the same accelerometer as that installed inside the blocks. The sampling rate of this accelerometers is 6 Hz. They were used for time synchronization of movement data between them and the accelerometer attached to the wrists.

2.3 Data Analysis

To visualize toy manipulations and determine behavioral imitation by using our blocks, we detected block motions and conducted correlation analysis. The signal processing to extract behavioral imitation consisted of four steps. First, the square sum of the three-axis acceleration data of the blocks were calculated. Second, block movements by the participants were detected from the instantaneous amplitude of the square sum of the three-axis acceleration data. Third, the manipulator (whether the block was moved by the manipulator or imitator) and the hand (whether the block was moved with the left or right hand) was distinguished through the correlation analysis between the block-movement data and the data of the manipulator's and imitator's hand movements. The hand that had the highest correlation coefficient with the block-movement data was determined as the hand that manipulated the block. Finally, behavioral imitation was determined by detecting two consecutive block movements of the manipulator and imitator. In addition, to determine the extent of behavioral imitations during parent-child interactions, the Pearson's correlation coefficients and delay times of the obtained behavioral imitations were evaluated through correlation analysis between block movements manipulated by the manipulator and imitator.

2.4 Results and Discussion

By using our blocks, movements of blocks and hands could be visualized (Fig. 2). For example, the manipulator moved Block A1 by using his left hand and the imitator subsequently moved Block B1 by using the same hand as the manipulator (blue and green arrows in Fig. 2). By visualizing such movements, behavioral imitations were visually confirmed. It should be also noted that there were periods when accelerometer data partially failed to be acquired (red dotted line in Fig. 2) because of connection failure between the accelerometers and visualization system. The success rate of data acquisition was 88.4%.

Behavioral imitations were extracted four times during correlation analysis. Their correlation coefficients ranged from 0.37 to 0.84 and delay times were 1 to 2 s (Table 1). They were consistent with the visually confirmed similarities and delay times by the visualization system of the digital blocks.

Movements of blocks and hands could also be visualized. Behavioral imitations and their correlation coefficients and delay times were obtained using our blocks. Thus, we confirmed the potential of visualization and quantitatively evaluation of behavioral imitation during toy play by using these blocks. For future work, the communication method between the accelerometers and visualization system should be improved to increase the success rate of data acquisition for robust determination of behavioral imitation. In addition, it should be useful to understand a relationship between the quality of behavioral imitation and the correlation coefficients and delay times with more parents and children pairs.

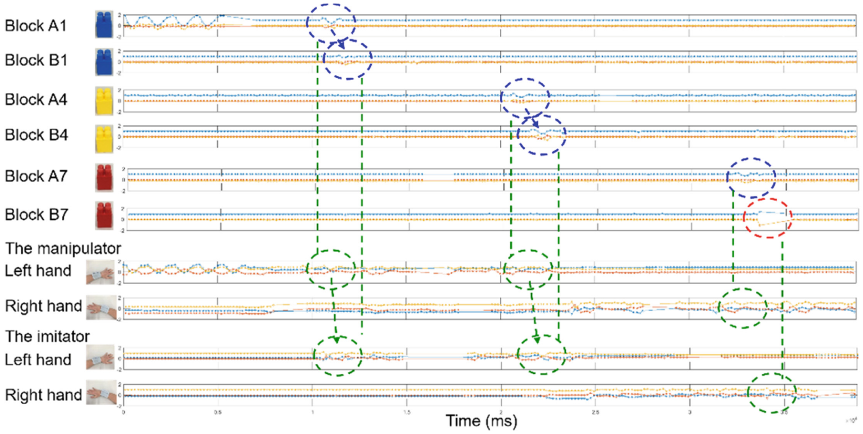


Fig. 2. Examples of three-axis acceleration data of blocks and hands during imitation task. Blue circles indicate movements of blocks. Green lines and circles indicate examples of movements of hands, and green arrows indicate behavioral imitations of block movements between manipulator and imitator. Red circles indicate lack of acceleration data. (Color figure online)

Table 1. Correlation coefficients and delay times between pairs of imitated–imitating blocks.

Block	Correlation coefficient	Delay time (ms)
Block A1–Block B1	0.37	1044
Block A4–Block B4	0.84	1134
Block A6–Block B6	0.81	1285
Block A8–Block B8	0.45	1610

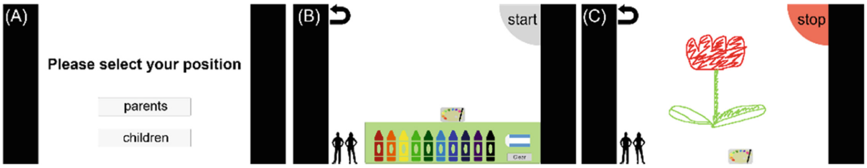
3 Development and Evaluation of Digital Drawing Application

3.1 Development of Digital Drawing Applications

Since there are many drawing applications available, parents and children can intuitively operate them to draw. Therefore, we developed a digital drawing application for visualization and determination of behavioral imitation during drawing by parents and children.

The drawing application was developed for Windows by means of Unity. It was designed to display a drawing screen after the role (parent or child) was selected on the top screen (Fig. 3(A)). On the drawing screen, the application enables the parent and child to select a color from 11 colors by selecting a crayon icon of that color (Fig. 3(B)), and to draw color lines (Fig. 3(C)).

Our drawing application enables us to acquire time-series data of the color and coordinates of the screen during drawing, which enabled us to develop a system to visualize drawing and obtain behavioral imitations between parents and children.



Please select your position

Fig. 3. Developed digital drawing application. (A) Initial screen, (B) main screen during color choice, and (C) example of digital drawing.

3.2 Experimental Procedure

In the experimental environment, two parent and child pairs who have been in PCIT were asked to sit side-by-side, and a tablet and digital pen (DTK-1660E, Wacom) were placed in front each of them. Data from the participants were obtained following receipt of written informed consent.

The evaluation experiment was conducted during the participants' PCIT sessions; our application was used in a real use case. As in normal PCIT, parents and children were asked to play freely by using our drawing application for 35 min. For the first five minutes, PCIT therapists observed behaviors of parents and children and coded interactions between them. After the first five minutes, parents and children were asked to continue free play while the parents were coached to improve their behavior including imitating the children's drawings for the next 30 min as therapy (Fig. 4).

To visualize drawings of parents and children and to determine the behavioral imitation between them, time-series data of colors and coordinates of the drawings were obtained. The sampling rate of the color measurement and coordinates was 30 Hz.

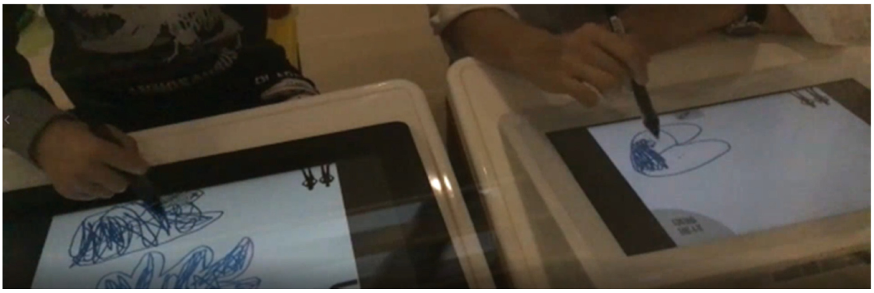


Fig. 4. Scene of experiment to evaluate digital drawing application.

3.3 Data Analysis

We visualized the drawings of parents and children and applied dynamic time warping (DTW) analysis to determine behavioral imitations [12]. DTW analysis enables us to calculate the similarity between two sets of time-series data that have different data lengths. This DTW analysis consisted of three steps. First, all combinations of Euclidean distances between each coordinate of the time-series drawing data of parents and children were calculated. Second, the combinations having minimum distances were derived. The combinations could overlap. Finally, the sum of the minimum distances was calculated as the DTW distance. The DTW distance was short when the drawing strokes were accurately imitated, and calculating the DTW distance would be useful to determine the extent of behavioral imitations during parent-child interactions.

3.4 Results and Discussion

By using our drawing application, colors and coordinates of the drawing of the parents and children could be visualized (Fig. 5). When the PCIT therapists considered that behavioral imitation between a parent and child was performed, colors and coordinates were visually overlapped in near time (dotted circles in Fig. 5(A)–(C)).

The DTW distance was derived during the behavioral imitation between a parent and child for drawing using blue (Fig. 5(D)). The DTW distance was relatively low at first then increased. This result indicates that the drawing strokes for the external shape were similar and that of filling in with blue was not similar between the drawings of the parent and child.

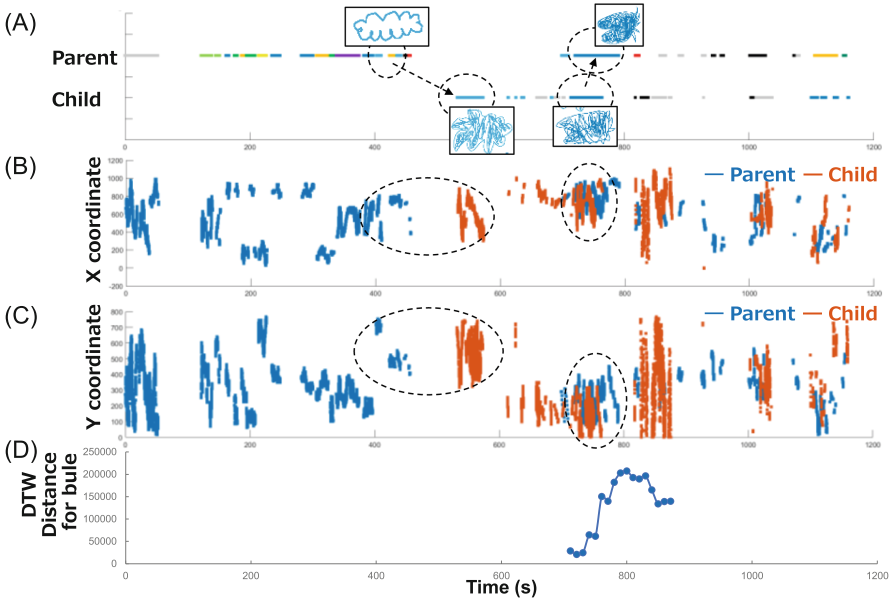


Fig. 5. Examples of digital drawing data obtained during PCIT by using our drawing application. (A) Time-series data of color, (B) those of x coordinates, (C) those of y coordinates, and (D) those of DTW distances for blue. (Color figure online)

From the results, drawings of parents and children were visualized, and behavioral imitations of drawings and their similarity were determined using our digital drawing application. Thus, we confirmed the evaluation possibility of visualization and quantitative evaluation of behavioral imitation during drawing by using our drawing application.

4 Conclusion

We developed two types of toys; blocks equipped with sensors and a digital drawing application. Manipulations could be visualized when playing and behavioral imitations between parents and children could be determined using these toys. We confirmed that the developed toys have the potential to extract the number and extent of behavioral imitations by correlation analysis or dynamic time warping analysis. Thus, they should be useful to evaluate parenting skills. For future work, we will fully automate signal processing for extracting behavioral imitation and verify our toys and signal processing with more parents and children pairs. This will be effective to develop practical and robust applications to evaluate behavioral imitation during parent-child interactions.

References

1. World Health Organization: Global status report on preventing violence against children 2020. World Health Organization Report, pp. 1–332 (2020)
2. Eyberg, S.M., Robinson, E.A.: Conduct problem behavior: standardization of a behavioral rating scale with adolescents. J. Clin. Child Psychol. **12**(3), 347–354 (1983)
3. Thomas, R., Zimmer-Gembeck, M.J.: Parent-child interaction therapy: an evidence-based treatment for child maltreatment. Child Maltreat. **17**(3), 253–266 (2012)
4. Kennedy, S.C., Kim, J.S., Tripodi, S.J., Brown, S.M., Gawdy, G.: Does parent-child interaction therapy reduce future physical abuse? A meta-analysis. Res. Soc. Work. Pract. **26**(2), 147–156 (2016)
5. Kawasaki, M., Kamo, T.: Internet-delivered parent-child interaction therapy (PCIT) in Japan: case report of application to a maltreating parent-child dyad. Arch. Clin. Med. Case Rep. **4**(6), 1218–1233 (2020)
6. Chartland, T.L., Bargh, J.A.: The chameleon effect: the perception-behavior link and social interaction. J. Pers. Soc. Psychol. **76**(6), 893–910 (1999)
7. Hale, J., Hamilton, A.F., De, C.: Cognitive mechanisms for responding to mimicry from others. Neurosci. Biobehav. Rev. **63**, 106–123 (2016)
8. Numata, T., et al.: Achieving affective human-virtual agent communication by enabling virtual agents to imitate positive expression. Sci. Rep. **10**, 5977 (2020)
9. Kato, K., Ienaga, N., Sugiura, Y.: Motion estimation of plush toys through detachable acceleration sensor module and machine learning. In: Stephanidis, C. (ed.) HCI International 2019 - Posters, vol. 1033, pp. 279–286. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-23528-4_39
10. Woodward, K., Kanjo, E., Brown, D.J., Inkster, B.: Tangtoys: smart toys that can communicate and improve children's wellbeing. In: Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers, pp. 497–499 (2020)
11. Wang, X., Takashima, K., Adachi, T., Finn, P., Sharlin, E., Kitamura, Y.: AssessBlocks: exploring toy block play features for assessing stress in young children after natural disasters. Proc. ACM Interact. Mobile Wearable Ubiquit. Technol. **4**(1), 1–29 (2020)
12. Keogh, E., Ratanamahatana, C.A.: Exact indexing of dynamic time warping. Knowl. Inf. Syst. **7**(3), 358–386 (2004). https://doi.org/10.1007/s10115-004-0154-9

Estimating Peer Evaluation Potential by Utilizing Learner Model During Group Work

Changhao Liang(✉), Thomas Gorham, Izumi Horikoshi,
Rwitajit Majumdar, and Hiroaki Ogata

Kyoto University, Kyoto 606-8501, Japan
liang.changhao.84c@st.kyoto-u.ac.jp

Abstract. Evaluation plays a substantial role in group work implementation and peer evaluation gets prevalent with increasing flipped learning scenarios and online evaluation platforms. The accuracy of peer evaluation remains contingent in group work practice thus eliciting relevant studies on grader reliability. In this study, we present a data-driven approach to solving this issue utilizing learner models. On the one hand, we use previous learning logs to estimate and visualize the grader reliability in group work evaluation sessions as "peer evaluation potential", which is used to align peer rating accuracy. On the other hand, leveraging reliability indicators created in the current session, learner models can be updated with new dimensions for subsequent usage. In addition, a case study in a high school English class was presented to examine this data-driven workflow and the results suggest the estimated peer evaluation potential correlates with the deviation from average peer judgment. Further potentials to cultivate peer evaluation-related capabilities are proposed as well.

Keywords: Peer evaluation · Peer assessment · Grader reliability · Learner model · Group work · CSCL

1 Introduction

Collaborative skill, as one of the critical soft skills nowadays, is consistently underscored in modern society. With increasing online lectures under the global pandemic, collaborative learning in small groups as group work gets prevalent for cultivating interpersonal skills [2]. In group work activities, evaluation is an indispensable component that can provide a score of group work outcome, give motivation, and promote learning with formative feedback [6].

As a critical tool for grading group work tasks, online peer evaluation tools have been widely adopted with increasing flipped learning practice [14]. However, its reliability remains questionable [11]. To compensate for this issue, enhancing the accuracy of peer ratings is highlighted in recent studies [2,10]. Before the cultivation of peer evaluation reliability, it is not a trivial thing to operationalize

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 287–294, 2022.
https://doi.org/10.1007/978-3-031-20218-6_20

this concept considering how to measure it in empirical practice and improve the data quality of peer evaluation. With the accumulation of immense online learning logs and the scaffold of learning analytics, we find an opportunity to model such capabilities in a data-driven environment.

In this study, we estimate and visualize the grader reliability in group work evaluation sessions as "peer evaluation potential". By examining how learner model indicators from a data-driven infrastructure can be utilized to estimate such coefficient, we employ it to align peer evaluation scores, and in turn, propose its potential for iterative use and update of learner models which leverage reliability indicators generated in peer evaluation sessions of group work.

This paper first introduces the research background of existing reliability estimation approaches and a data-driven environment. Then we put forward our learner model-based solution to estimate peer evaluation potential followed by a preliminary case study in an authentic classroom.

2 Research Background

2.1 Peer Evaluation and Its Reliability

Peer evaluation activity can be interpreted as the process of assessing the work of peers against a set of various assessment criteria [12]. In the existing group work practice, the evaluation from only the teacher's perspective is limited since one teacher cannot identify the performance of every participant [7] and social loafing and free riding phenomenon exist [13]. Peer evaluation is required to alleviate teachers' workload and provide an overall inspection across the group work process [3,15].

In parallel, peer evaluation reliability was underscored since the quality of peer evaluation remains promising [2]. Current studies in online environments present several scaffolds to improve the reliability of peer assessment with enhanced privacy [14] and group awareness support [7]. Nevertheless, there remains an unbalance of grader reliability due to individual difference among learners, which lead to less accurate evaluation results in practice. To address this issue, researchers made attempts to alter the final rating values according to grader-specific variables such as previous rating tendency [10] and previous grades of relevant tasks [3,11]. However, the possibilities of learning model data are seldom explored to make a comprehensive prospect of these variables with existing peer evaluation designs holding limited data aggregation and re-use features [9].

2.2 Learner Model and GLOBE Framework

In learning science, researchers conduct learner modeling where the learning evidence data is divided into domain-specific and domain-independent information as quantified indicators [4]. These indicators cover recorded learning behaviors on the learning management platforms (LMS), preferred learning styles, previous group working experiences, and so on [5].

To make the best of learner modal data, Group Learning Orchestration Based on Evidence (GLOBE) was put forward to support group work implementation in the data-driven context [8]. GLOBE consists of four phases with corresponding systems for data processing: group formation, orchestration, evaluation, and reflection. The peer evaluation system [9] provides a data-driven infrastructure for further research on peer evaluation behaviors and capability cultivation.

In former studies with GLOBE, learner model indicators from previous group work evidence were used to form groups heterogeneously or homogeneously like [1]. In the evaluation phase, these indicators can be used to estimate peer evaluation potentials as well [11]. Similar to the data-driven group formation implementation underpinned by iterative GLOBE data circulation [8], evaluation functionalities will create new evidence in the data circulation to update the learner model, which can promote fine-tuning of the accuracy of peer evaluation results with estimated grader reliability for following sessions.

3 Methodology

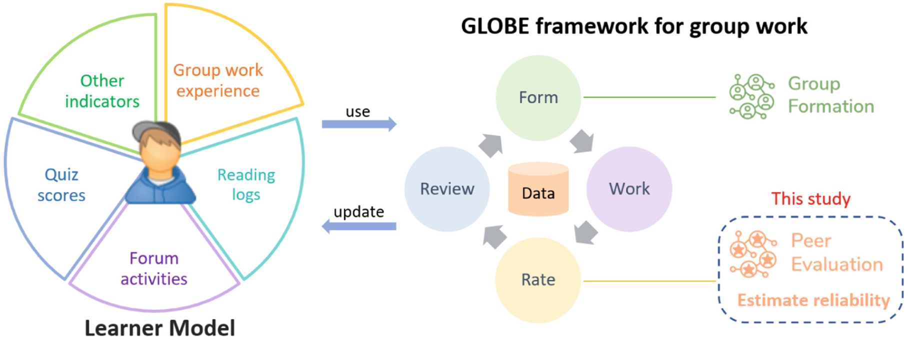


Fig. 1. Iterative use and update of learner model with group work implementations under GLOBE

Figure 1 shows the iterative use and update of learner models with group work implementations under GLOBE. The learner modeling for grader reliability in this study follows this workflow. In the first step, the peer evaluation potential coefficient is estimated by learner model data emanating from learning evidence from various dimensions. This potential coefficient is used to weigh the scores given by each grader in the current peer evaluation activity, and the teacher can check the aligned score of each participant. Meanwhile, other indicators of evaluation reliability emanating from the peer evaluation engagement and dynamic score distribution are calculated synchronously during the proceeding peer evaluation session. These indicators of grader reliability, along with group work experience as weighed rating scores of each evaluatee, will converge into the GLOBE data stream as updated learner models for subsequent activities.

3.1 Estimating Peer Evaluation Potential from Learner Model Data

In the GLOBE systems, learners are embodied as vectors featured by different learner model indicators such as reading attributes and quiz scores. To estimate the peer evaluation potential coefficient (P), we adopt the same learner model indicators used in the group formation stage, since the group work and its corresponding peer evaluation activity has a shared learning context. The P value of one grader is calculated from n quantified learner model characteristics (c) according to (1) and standardized into than range of 0 to 1. In the beginning, each indicator (i) has an equal effect (w) which could be adjusted in practice. This peer evaluation potential coefficient will be used as the weight of each grader (g) in the weighted average calculation of the final scores (S) from m graders' raw rating scores (r). Even though there is no existing learner model data, instructors can start from group work with random grouping using GLOBE systems, then the group work data such as the final scores (S) will end up in the learner model and be available in the next round.

$$P = \frac{\sum w_i c_i}{n}, S = \frac{\sum P_g r_g}{m} \tag{1}$$

3.2 Calculating Peer Evaluation Indicators of Current Activity

Besides the peer evaluation potential estimated from previous learning logs, other indicators from the current peer evaluation activity reflecting the grader reliability are considered as well. For peer evaluation of other groups, the deviation from the teacher's rating (r_t) and that from the mean of all peer ratings ($\bar{r}$) are meaningful to model the grader reliability. Accordingly, we created rating validity (V) and rating reliability (R) of each grader underpinned by (2) from the distribution of their rating values of G group candidates (c), with a lower value indicates higher grader reliability. The values of V and R to each candidate group are dynamically visualized to the teacher during the evaluation activity and can update the learner model of group work experience after standardization by (3). Extreme values that exceed two standard deviations of the mean will be excluded before implementing (3) with a zero value.

$$V = \frac{\sum |r_{t_c} - r_c|}{G}, R = \frac{\sum |\bar{r}_c - r_c|}{G} \tag{2}$$

$$V_{std} = \frac{1 - V}{V_{max}}, R_{std} = \frac{1 - R}{R_{max}} \tag{3}$$

3.3 Updating Learner Model with Peer Evaluation Indicators for Next Round Activity

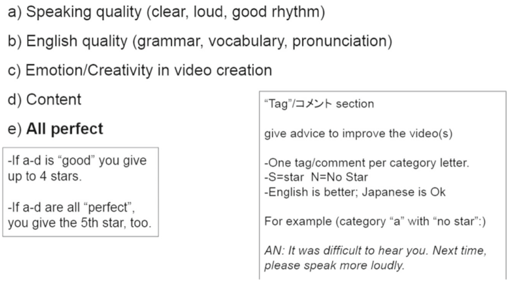
After the peer evaluation session, the P value and final V, R values can be snapshotted as learner models together with all rating scores for the next round of group work or other learning analytics tools. These indicators will be iteratively used to determine the P value of the corresponding grader in the following peer evaluation as c in (1), with a certain weight along with indicators of other learner model indicators.

In the next step, more evaluation reliability-related indicators from higher-level modeling capitalized on semantic analysis of textual feedback and behavior evidence will be integrated into the snapshot. Since these indicators are dynamically updated with the process of group evaluation submissions from the teacher and students, they cannot be used as predictors for reliability estimation of the current group work but can enrich the learner model for subsequent group work evaluation sessions with similar learning contexts.

4 Case Study

To walk through the modeling workflow and preliminarily examine the effect of the estimation, data from the peer evaluation sessions of small-group English movie presentations from 18 high school students was analyzed in this case. During the peer evaluation activity, each student was required to give ratings and feedback according to rubrics given by the teacher (see Fig. 2). The peer evaluation activities were conducted for two rounds and there were two students absent from each round respectively. For each round, heterogeneous groups of 3 students were formed based on their previous English test scores, which were used for estimation of the P value as well. As is shown in Fig. 3, the teacher can examine and store P values with adjusted peer ratings score for group members, V, R indicators of grader reliability calculated from rating data (see Table 1), and textual feedback during the peer evaluation session in a real-time manner. Self-evaluation scores are also displayed in the panel, though they are not addressed in the current modeling process. In the second round, the P value was updated employing standardized V and R indicators (3) of the first round as c in (1) with equal w of English test score indicators. The descriptive statistics for P, V, R are given in Table 2, and correlation analysis was conducted among these learner model attributes for grader reliability to inspect the power of P (see Fig. 4). One student with V and R values exceeding two standard deviations of mean was excluded from the correlation analysis.

As for the results, negative relation is detected between P and R in the first round, connoting correspondence of peer evaluation potential coefficient to rating reliability which conforms to [11]. However, the positive relation between P and V reflects a disparity in the teacher's rating and so does that between R and V. After confirmation with the teacher, we inferred that it was because the teacher tended to give higher scores for encouragement for the first group work. This can also explain the inconsistency between V and R in the first round. The correlation between P and V, P and R in the second round activity gets obscure, which can be caused by the instability of the modeling with limited iteration and small sample size. V and R are closely correlated in the second round, which is expected since both of them depict agreement with others.



a) Speaking quality (clear, loud, good rhythm)

b) English quality (grammar, vocabulary, pronunciation)

c) Emotion/Creativity in video creation

d) Content

e) **All perfect**

-If a-d is "good" you give up to 4 stars.

-If a-d are all "perfect", you give the 5th star, too.

"Tag"/コメント section

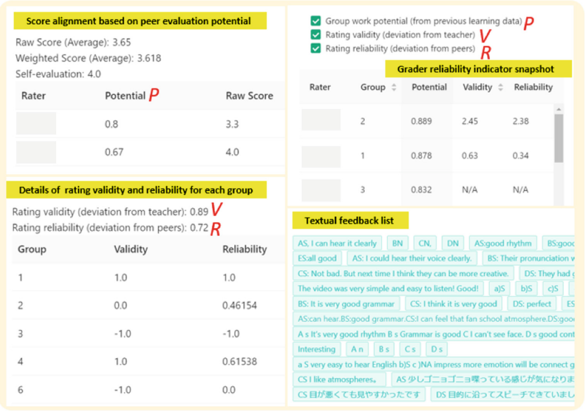
give advice to improve the video(s)

-One tag/comment per category letter.
-S=star N=No Star
-English is better; Japanese is Ok

For example (category "a" with "no star":)

AN: It was difficult to hear you. Next time, please speak more loudly.

Fig. 2. Rubrics for the peer evaluation session of English movie presentations.



Score alignment based on peer evaluation potential

Raw Score (Average): 3.65
Weighted Score (Average): 3.618
Self-evaluation: 4.0

Rater	Potential P	Raw Score
	0.8	3.3
	0.67	4.0

Details of rating validity and reliability for each group

Rating validity (deviation from teacher): 0.89 V
Rating reliability (deviation from peers): 0.72 R

Group	Validity	Reliability
1	1.0	1.0
2	0.0	0.46154
3	-1.0	-1.0
4	1.0	0.61538
6	-1.0	0.0

☑ Group work potential (from previous learning data) P
☑ Rating validity (deviation from teacher) V
☑ Rating reliability (deviation from peers) R

Grader reliability indicator snapshot

Rater	Group	Potential	Validity	Reliability
	2	0.889	2.45	2.38
	1	0.878	0.63	0.34
	3	0.832	N/A	N/A

Textual feedback list

AS, I can hear it clearly | BN | CN, | DN | AS:good rhythm | BS:gooc
ES:all good | AS: I could hear their voice clearly. | BS: Their pronunciation w
CS: Not bad. But next time I think they can be more creative. | DS: They had g
The video was very simple and easy to listen! Good! | a)S | b)S | c)S
BS: It is very good grammar | CS: I think it is very good | DS: perfect | ES
AS:can hear.BS:good grammar.CS:I can feel that fan school atmosphere.DS:gooc
A s It's very good rhythm B s Grammar is good C I can't see face. D s good cont
Interesting | A n | B s | C s | D s
a S very easy to hear English b)S c)NA impress more emotion will be connect g
CS I like atmospheres。 | AS 少しゴニョゴニョ喋っている感じが気になりま
CS 目が悪くても見やすかったです | DS 目的に沿ってスピーチできていまし

Fig. 3. Interfaces of peer evaluation attributes visualization

The case study provided an example to apply the learner models in the classroom peer evaluation activity. Beyond this case, the proposed approach can be also employed in other contexts such as online courses. To overcome the limitations, further studies with more participants and iterative peer evaluation sessions are required to assess the performance advantages of this learner model-driven approach. Since this is a preliminary study, the weight of the grader reliability indicators (w) should be re-considered in more trials as well.

Table 1. Peer and teacher's ratings for groups to calculate V and R

Trial	Learner model attribute	N	Mean	SD	Min	Max
First round	Peer rating for groups (r) [1–5]	80	3.75	0.893	1.000	5.000
	Teacher's rating for groups (r_t) [1–5]	6	4.33	0.516	4.000	5.000
Second round	Peer rating for groups (r) [1–5]	80	4.088	1.105	1.000	5.000
	Teacher's rating for groups (r_t) [1–5]	6	4.167	0.753	3.000	5.000

Table 2. Descriptive Statistics of learner model attributes for grader reliability

Trial	Learner model attribute	N	Mean	SD	Min	Max
First round	P [0-1]	16	0.667	0.148	0.461	0.889
	V [0-...]	16	1.086	0.303	0.630	1.600
	R [0-...]	16	0.836	0.288	0.320	1.370
Second round	P [0-1]	16	0.529	0.101	0.368	0.694
	V [0-...]	16	1.108	0.509	0.450	2.450
	R [0-...]	16	0.907	0.489	0.340	2.380

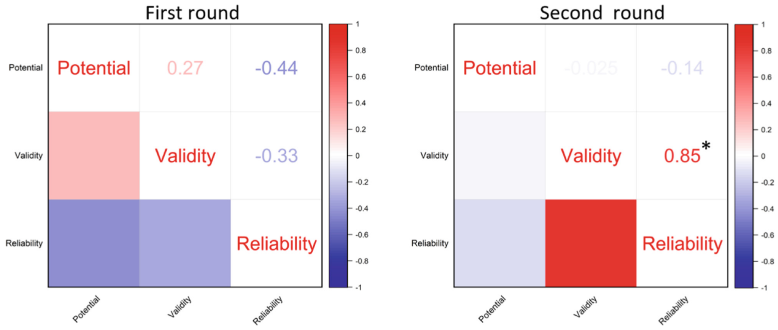


Fig. 4. Correlation analysis of peer evaluation potential and grader reliability indicators

5 Conclusion and Future Work

In this paper, we introduced a data-driven approach to estimating peer evaluation potential coefficients from previous learner model indicators, which aims to enhance the grader reliability of peer evaluation in small-group collaborative learning. The learner model-based solution can enhance the accuracy of peer rating scores and visualize rating quality indicators for teachers in a real-time manner, hence lowering the threshold to conduct peer evaluation activities. A case study showed the actual process of modeling in a high school English class and the results suggest the estimated peer evaluation potential correlates with the deviation from average peer judgment.

The study also contributes to data accumulation for learner modeling via iterative updated grader reliability data from authentic peer evaluation activities, which can be used in learning analytics tools in the data-driven ecosystem. Compared to other solutions to model rater reliability, we make a comprehensive consideration under an LA-enhanced architecture with dynamically updated historical records of learner data in the calculation of evaluation quality attributes.

The estimation of peer evaluation reliability is just the first step toward the ultimate goal of cultivating peer evaluation-related capabilities. In the following study, we should repeat the learner model iteration process with a larger sample and more iterative trials, and consider how to present these peer evaluation reliability indicators as a feedback dashboard to students in a proper way, which can lead to the improvement of their peer evaluation skills.

Acknowledgements. This research was supported by JSPS KAKENHI 20K20131, 20H01722, 22H03902, NEDO JPNP18013, JPNP20006, and JST SPRING JPMJSP2110.

References

1. Abou-Khalil, V., Ogata, H.: Homogeneous student engagement: a strategy for group formation during online learning. In: Hernández-Leo, D., Hishiyama, R., Zurita, G., Weyers, B., Nolte, A., Ogata, H. (eds.) CollabTech 2021. LNCS, vol. 12856, pp. 85–92. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-85071-5_6
2. Aminu, N., Hamdan, M., Russell, C.: Accuracy of self-evaluation in a peer-learning environment: an analysis of a group learning model. SN Soc. Sci. **1**(7), 1–17 (2021)
3. Bjelobaba, G., Paunovic, M., Savic, A., Stefanovic, H., Doganjic, J., Miladinovic Bogavac, Z.: Blockchain technologies and digitalization in function of student work evaluation. Sustainability **14**(9), 5333 (2022)
4. Boticki, I., Akçapınar, G., Ogata, H.: E-book user modelling through learning analytics: the case of learner engagement and reading styles. Interact. Learn. Environ. **27**(5–6), 754–765 (2019)
5. Bozic, N.H., Mornar, V., Boticki, I.: Introducing adaptivity and collaborative support into a web-based LMS. Comput. Inform. **27**(4), 639–659 (2008)
6. Forsell, J., Forslund Frykedal, K., Hammar Chiriac, E.: Group work assessment: assessing social skills at group level. Small Group Res. **51**(1), 87–124 (2020)
7. Kasch, J., van Rosmalen, P., Löhr, A., Klemke, R., Antonaci, A., Kalz, M.: Students' perceptions of the peer-feedback experience in MOOCs. Distance Educ. **42**(1), 145–163 (2021)
8. Liang, C., Majumdar, R., Ogata, H.: Learning log-based automatic group formation: system design and classroom implementation study. Res. Pract. Technol. Enhanced Learn. **16**(1), 1–22 (2021). https://doi.org/10.1186/s41039-021-00156-w
9. Changhao, L., Toyokawa, Y., Nakanishi, T., Majumdar, R., Ogata, H.: Supporting peer evaluation in a data-driven group learning environment. In: Hernández-Leo, D., Hishiyama, R., Zurita, G., Weyers, B., Nolte, A., Ogata, H. (eds.) CollabTech 2021. LNCS, vol. 12856, pp. 93–100. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-85071-5_7
10. Masaki, U., Maomi, U., et al.: Item response theory with assessors' parameters of peer assessment. J. Inst. Electron. Inf. Commun. Eng. **91**(2), 377–388 (2008)
11. Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., Koller, D.: Tuned models of peer assessment in MOOCs. arXiv preprint arXiv:1307.2579 (2013)
12. Pond, K., Ul-Haq, R.: Learning to assess students using peer review. Stud. Educ. Eval. **24**, 331–348 (1997)
13. Strijbos, J.W.: Assessment of (computer-supported) collaborative learning. IEEE Trans. Learn. Technol. **4**(1), 59–73 (2010)
14. Ismail, N., et al.: Peer evaluation system in team work skills assessment. In: Fook, C.Y., Sidhu, G.K., Narasuman, S., Fong, L.L., Abdul Rahman, S.B. (eds.) 7th International Conference on University Learning and Teaching (InCULT 2014) Proceedings, pp. 603–616. Springer, Singapore (2016). https://doi.org/10.1007/978-981-287-664-5_47
15. Willey, K., Gardner, A.: Investigating the capacity of self and peer assessment activities to engage students and promote learning. Eur. J. Eng. Educ. **35**(4), 429–443 (2010)

Scenario for Analysing Student Interactions and Orchestration Load in Collaborative and Hybrid Learning Environments

Adrián Carruana Martín[1](✉), Alejandro Ortega-Arranz[2], Carlos Alario-Hoyos[1], Ishari Amarasinghe[3], Davinia Hernández-Leo[3], and Carlos Delgado Kloos[1]

[1] Universidad Carlos III de Madrid, Leganes, Spain
acarruan@inf.uc3m.es, {calario,cdk}@it.uc3m.es
[2] Universidad de Valladolid, Valladolid, Spain
alex@gsic.uva.es
[3] Universitat Pompeu Fabra, Barcelona, Spain
{ishari.amarasinghe,davinia.hernandez-leo}@upf.edu

Abstract. Educational environments have been affected by the COVID-19 pandemic and have evolved to support classes, which involve in some cases synchronous hybrid learning environments. These environments enable students attend classes online and on-site simultaneously. Synchronous hybrid environments provide a greater flexibility for students but, in contrast, are likely to increase teachers' orchestration load and decrease interactions between students, especially between those online and those on-site. This study proposes a scenario to explore the factors affecting the orchestration load and the student interactions in collaborative and synchronous hybrid learning environments. The scenario involves the use of a collaborative learning flow pattern (jigsaw) and the technologies that will enable the data collection to understand such factors affecting to orchestration load and interaction. The outcomes from the implementation of this scenario will provide useful insights to further understand the benefits and limitations of synchronous hybrid learning environments.

Keywords: Hybrid learning · Collaborative learning · Teacher orchestration · Scenario design · Teacher agency

1 Introduction

Many current educational environments have been affected by the COVID-19 pandemic, including higher education [1]. Social distancing measures aimed at a reduction of students per classroom lead to hybrid environments (e.g., use of mirror classrooms, some students at home, etc.) [2]. Some of these environments still remain nowadays, and are seen as an alternative to traditional environments.

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 295–303, 2022.
https://doi.org/10.1007/978-3-031-20218-6_21

International organisations such as UNESCO are stressing the importance of these hybrid environments in the current society [3].

These hybrid learning environments provide a higher flexibility for students, as they allow participate in class from anywhere. However, these environments are likely to increase the orchestration load of teachers during collaborative tasks, as they require to manage on-site and online students and the questions-requests arising from both on-site and hybrid modes [4]. Also, these environments are likely to decrease interactions between students as compared to other traditional on-site environments (e.g., physical classrooms), as the use of a tool for the communication between students can cause difficulties [2]. This transition from on-site to synchronous hybrid learning can have an impact on teachers, more specifically on their agency [2]. For these reasons, this work-in-progress paper presents a scenario from which useful data can be collected in order to achieve the following research goals:

1. To extract the factors contributing to orchestration load and to explore students' interactions in synchronous hybrid computer-supported collaborative learning (SH-CSCL) environments.
2. To extract the factors contributing to teachers' agency in SH-CSCL environments.

2 Related Works

There are not any scenarios that have explored all three characteristics sought here at the same time (orchestration load, teacher agency or synchronous hybrid learning environments), not even two of them. For that reason, relevant papers including at least one of the features are following described. Moreover, as the focus of the scenario is on higher education, the studies presented in this section do so as well.

One of the papers on synchronous hybrid learning environments is the one by Bülow [5]. In this paper different hybrid spaces, including synchronous ones, are analysed. After the analysis, the author specifies the factors that should be covered in hybrid learning environments to make them efficient. The main characteristics to be fulfilled are: good group formation must be carried out and collaboration between students must be facilitated. Another paper dealing with synchronous hybrid learning environments is the one by Flynn-Wilson & Reynolds [7]. In this paper, classes are conducted in hybrid environments, asynchronous and synchronous, during four semesters. The authors concluded that students preferred the synchronous learning environment but that more care had to be taken to avoid technical failures (Internet downtime, microphone failure, etc.).

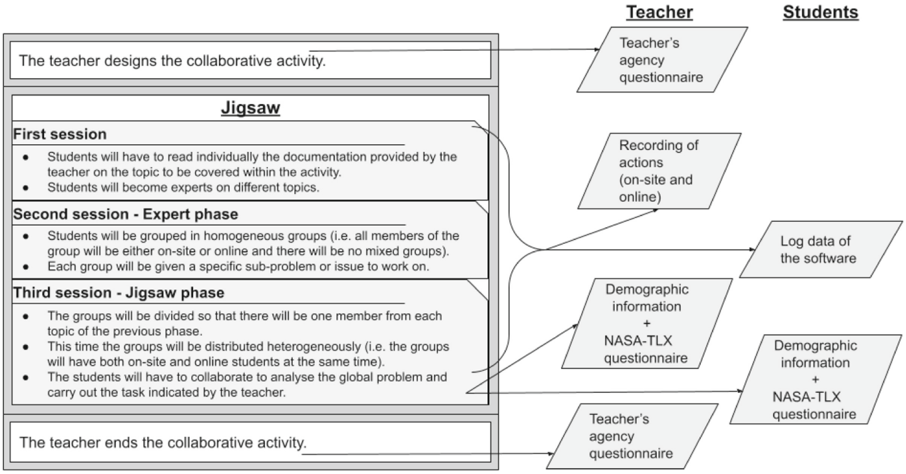
Among the most famous works about orchestration load is is the one by Prieto et al. [8]. In this paper, orchestration with the use of a tool for collaboration is evaluated. From this evaluation, the authors conclude that there is a big difference in the orchestration load in the first uses of the application than after a regular use. Another paper on orchestration load is that of Wang et al. [9]. This paper evaluates different factors of orchestration according to the teacher's use of the educational software such as the number of times the audio is played, number of corrections, etc. One of the conclusions was that performing a new task increases the orchestration load.

Finally, regarding previous works about teachers' agency, it is worth highlighting that by Sammons et al. [10]. In this paper, several scenarios for the analysis of variations in teachers' work and life are carried out. The authors conclude that the fewer changes in a teacher's work, the higher the teacher's agency. Another paper on teacher's agency is that of Kayi-Aydar [11]. In this paper, the teacher's agency is analysed according to his or her position in solving a problem. The author comes to the conclusion that what most affects agency is the context in which teachers find themselves when solving problems.

3 Description of the Scenario

This work-in-progress paper aims to show a hybrid collaborative learning situation design with which to obtain relevant data to address the research questions mentioned above.

The scenario requires software capable of transmitting the synchronous sessions to online students and capable of receiving and answering questions and doubts from all students. In addition, the scenario can involve as few as 9 students, the recommended minimum to be able to perform the jigsaw pattern correctly [6], and as many as 90 students so that the experience is not too large [12]. The proposed scenario can be divided into three consecutive parts in time, as shown in Fig. 1. Furthermore this figure shows the data sources used.


Teacher
Students
Teacher's agency questionnaire
Recording of actions (on-site and online)
Log data of the software
Demographic information + NASA-TLX questionnaire
Demographic information + NASA-TLX questionnaire
Teacher's agency questionnaire
Jigsaw
First session
Second session - Expert phase
Third session - Jigsaw phase

The teacher designs the collaborative activity.

- Students will have to read individually the documentation provided by the teacher on the topic to be covered within the activity.
- Students will become experts on different topics.

- Students will be grouped in homogeneous groups (i.e. all members of the group will be either on-site or online and there will be no mixed groups).
- Each group will be given a specific sub-problem or issue to work on.

- The groups will be divided so that there will be one member from each topic of the previous phase.
- This time the groups will be distributed heterogeneously (i.e. the groups will have both on-site and online students at the same time).
- The students will have to collaborate to analyse the global problem and carry out the task indicated by the teacher.

The teacher ends the collaborative activity.

Fig. 1. Scenario and data sources.

The first part of the scenario, corresponding to the research part, is the pre-activity questionnaire for teachers. Before any activity is carried out and even before planning them, teachers should take a questionnaire. This questionnaire, based on the one proposed by Hull et al. [13], will capture the state of the agency prior to the new collaborative activity in synchronous hybrid learning environments.

The second part of the scenario, corresponding to the pedagogical part, is the collaborative learning activity. In this activity a CLFP is used, more specifically the jigsaw pattern [6]. This CLFP consists of three different phases, carried out on different sessions. The first phase consists on individually, students have to read the documentation given by the teacher about a concrete topic. Students become experts on different topics. The second phase is the so-called expert phase. In this phase, students are grouped in homogeneous groups (i.e., all members of the group are either on-site or online and there are no mixed groups). Each group is assigned a specific sub-problem or issue to work on. The third phase is the jigsaw phase. In this phase, the groups are divided so that there is one member from each topic of the previous phase. This time the groups are distributed heterogeneously (i.e., the groups have both on-site and online students at the same time). In this phase, the students have to collaborate to analyse the global problem and complete a questionnaire whose answers can be obtained from the previous documentation.

The third part of the scenario, corresponding to the research part, is the post-activity questionnaire for teachers. After the end of all the educational sessions, the teachers retake the questionnaire from the first part. This is done to be able to analyse the impact of these activities on the teachers' agency, to observe where the changes occur and in which characteristics. The values obtained are compared with the initial values to check whether some of the agency characteristics, such as teacher independence, have decreased or increased. The objective is to measure with these values the impact generated by collaborative activity in synchronous hybrid learning environments using software that supports these environments. Finally, teachers have to complete questionnaires to measure their orchestration load in the implementation of the learning situation and their demographic information. For the orchestration load, the NASA-TLX (Task Load Index) methodology questionnaire [14] is used. This questionnaire has been used in other studies of orchestration load in educational settings [15]. Students have to fill in questionnaires with demographic information and their experience in the activity.

The data sources from which the information for the objectives of the study is extracted must be taken into account for the scenario. One of the sources, in addition to the questionnaires discussed above, is the log data of the software used for the synchronous hybrid learning environments. From these records, the attendance of students, the number of times students talk, chat, etc. at group and class level are obtained. This data are used to measure the interaction and collaboration of the students during the sessions. The next source of data is the monitoring of students' work. This can be obtained through forms that students have to fill in after each phase of the CLFP. These forms allow us to know the final result of the collaboration between the students. The final source of data for the scenario is the recording of digital sessions and the collection of data by observers of the teacher's activities. The recordings of the digital sessions show the number and timing of teachers' interactions with the on-site and offsite students. The observations of the teachers' actions also help triangulate the orchestration load information gathered through the questionnaires. Discourse among teachers and students can be modelled using Epistemic Network Analysis (ENA) [16,17]. ENA aids in visualising the structure of connections among codes in discourse data using dynamic network models [16]. ENA has been applied in previous studies to model teachers' orchestration actions [18,19].

The decision to extract these data is related to the factors that are the objectives of this study. This relationship can be seen in Fig. 2. The source of data feeds not only the factors but also other sources. This happens with the teacher's records, obtained from the observer, together with the records obtained from the software used to make the dataset for the ENA [16,17]. Other information can also be obtained from the teacher's record that can be extracted with experts to try to obtain more information about the orchestration load. The software record is also used to extract the interactions of the students. For this purpose, the number of times students have communicated in the group or with the teacher, the time they have been talking, the use of the different options provided by the software, etc. are extracted. There are also questionnaires such as the NASA [14] and Agency [13]. Both have been used in other works [15] in the educational field with the same objectives, obtaining good results. Finally, there is the demographic information that serves as support for obtaining patterns in the information obtained, for example, if the Agency is different according to the age group of the teacher, if final year students are more active or use the software more easily, etc.

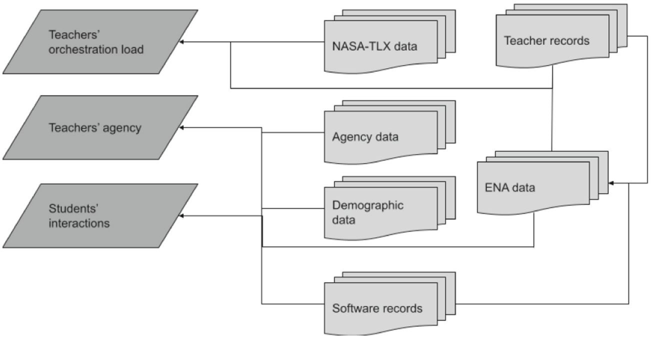


Fig. 2. Relationship between the extracted data and the objective factors.

Finally, an example of what the complete scenario could look like in a real environment is shown to clarify the different parts: "Marimar is a web developer teacher who carries out a collaborative activity with the jigsaw pattern in her class of 60 students (half of which are onsite and half online). Before designing the activity she carries out the questionnaire of the agency. Then, she designs the three phases of the jigsaw pattern. In the first phase, the teacher gives information to 20 students about JavaScript, 20 students about HTML and 20 students about PHP to read. In the second phase, she makes 10 groups with the students who have read about the same topic. Half of the groups are entirely on-site and the other half online. All students have to be connected to Engageli software and have to make a summary of their topic in a shared Google Docs document. In the last phase, students are again divided into 10 groups but in each group, there are two students from each topic. These groups are heterogeneous, some students are online and others on-site. The students work through Engageli. As an activity, they have to analyse the code of a small news web portal. In the end, the students do a quiz on certain features of the portal related to each of the topics and among the group members. They have to discuss which is the most appropriate solution. In addition, along with the answers, students have to fill in a questionnaire with their demographic information. During all the sessions an observer will take notes about the actions that the teacher carried out and the time needed to complete them, as well as which group did them if any were involved. Once all collaborative activities are finished, Marimar fills in the NASA-TLX questionnaire, the agency questionnaire and one to fill in her demographic information".

4 Preliminary Assessment

An implementation of the previous research design has been performed to collect some preliminary data. The learning situation involved a workshop on human-centered design with 17 students and 1 teacher. Most students in this evaluation master the topic of this activity but lack previous experience in hybrid learning environments. The students consisted of 8 PhD students and 9 PhD. Most of the students in this assessment are aged between 21 and 30.

Regarding the students' results of subscales on the orchestration load (according to the NASA-TXL questionnaire, the lower the better and between 30–40 would be considered a normal load [14]), it can be seen in Table 1. The values of the mental demand are a little high due, according to the students, to the difficulty of coordinating with their peers. Regarding the physical demand, the highest values are due, according to the students, to the large amount of noise during the activity. Regarding the time demand, the students justified their high values because they had technical problems that caused several delays, as indicated in the related literature. The performance is good, even with 1 student indicating the minimum value, although the students who had higher values justified it by lack of time. Effort is in line with the values of mental demand. The frustration level is good and the few students who had higher values were due to the stress of lack of time. Finally, the mean weighted workload score (final value of the NASA-TXL, calculated taking into account the weights of each subscale

[14]) of the students is 50.66, its median is 50.67 and the standard deviation is 12.21. These results were compared with a similar on-site collaborative study testing different collaboration strategies [20]. The orchestration load in our scenario was lower compared to all strategies used by the other study. More data needs to be collected as this difference may be due to the fact that the students in our workshop were more knowledgeable than in the compared studies [20].

Table 1. Orchestration load subscales of students

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	Mean	Std.Dev
Mental demand	50	60	10	60	80	67	75	60	50	35	70	80	30	60	70	70	60	58.06	18.6
Physical demand	40	40	1	90	70	8	60	20	20	10	30	10	60	5	33	1	20	30.47	26.33
Temporal demand	85	80	10	10	80	79	40	80	40	45	60	80	50	65	70	75	30	57.59	24.83
Performance	1	10	70	20	30	27	35	10	10	45	20	30	5	30	40	15	20	24.53	17.14
Effort	60	70	10	80	90	58	65	60	70	55	50	70	55	70	70	70	50	61.94	16.98
Frustration level	35	20	1	1	60	7	40	20	1	60	10	30	25	40	65	75	30	30.53	23.73
Workload	50	51	21	61	65	58	52	53	39	49	48	68	32	49	59	66	43	50.66	12.21

This workshop was coordinated together with the teacher due to time constraints. The teacher indicated that this joint coordination affected her agency and her orchestration load, so her results differed from those she would have had in a more accurate recreation of the scenario. These results for the teacher's orchestration load (according to the NASA-TXL questionnaire, the lower the better and between 30–40 would be considered a normal load) are as follows: Mental Demand: 50, Physical Demand: 20, Temporal Demand: 70, Performance: 60, Effort: 60 and Frustration Level: 60. The weighted workload score (final value of the NASA-TXL, calculated taking into account the weights of each subscale [14]) of the teacher is 57.34. The teacher's workload value is similar to that of the students. Although it is possible that they would have been higher if it were not for the joint coordination. On the other hand, the results of the teacher's agency show no change from before and after the scenario. This may be due, as the teacher commented, to our interference in the organisation of the workshop or because the activity simply did not interfere with her agency. More data will need to be collected to verify this.

5 Future Work

First, new methods for the analysis of orchestration and teacher agency will be explored. Of these new methods, special attention will be paid to qualitative methods, as the aim is to delve deeper into the different processes that the teacher performs in synchronous hybrid learning environments.

The next step is to redefine the proposed scenario with these new evaluation methods and incorporating improvements to avoid problems encountered in the small implementation that was done. Some of the main problems were the noise generated when working in groups, the technical problems that delayed the implementation and shortened the time, or the little freedom the teacher had to design and organise the activity.

Acknowledgements. This work was supported in part by grant PID2020-112584RB-C31, PID2020-112584RB-C32 and PID2020-112584RB-C33 funded by MCIN/ AEI /10.13039/501100011033, and in part by the Madrid Regional Government through the e-Madrid-CM Project under Grant S2018/TCS-4307 and under the Multiannual Agreement with UC3M in the line of Excellence of University Professors (EPUC3M21), and in the context of the V PRICIT (Regional Programme of Research and Technological Innovation), a project which is co-funded by the European Structural Funds (FSE and FEDER). Partial support has also been received from the European Commission through Erasmus+ Capacity Building in the Field of Higher Education projects, more specifically through projects InnovaT and PROF-XXI (598758-EPP-1-2018-1-AT-EPPKA2-CBHE-JP), (609767-EPP-1-2019-1-ES-EPPKA2-CBHE-JP). This publication reflects the views only of the authors and funders cannot be held responsible for any use which may be made of the information contained therein.

References

1. Cahapay, M.B.: Rethinking education in the new normal post-COVID-19 era: a curriculum studies perspective. Aquademia **4**(2), ep20018 (2020)
2. Raes, A., et al.: A systematic literature review on synchronous hybrid learning: gaps identified. Learn. Environ. Res. **23**(3), 269–290 (2020)
3. UNESCO Education in a post-COVID world: nine ideas for public actions (2020). https://en.unesco.org/sites/default/files/education_in_a_post-covid_world-nine_ideas_for_public_action.pdf
4. Prieto, L.P., et al.: Orchestration load indicators and patterns: in-the-wild studies using mobile eye-tracking. IEEE TLT **11**(2), 216–229 (2018)
5. Bülow, M.W.: Designing synchronous hybrid learning spaces: challenges and opportunities. Hybrid Learn. Spaces, 135–163 (2022). https://doi.org/10.1007/978-3-030-88520-5_9
6. Hernández-Leo, D., et al.: COLLAGE: a collaborative learning design editor based on patterns. ET&S **9**(1), 58–71 (2006)
7. Flynn-Wilson, L., Reynolds, K.E.: Student responses to virtual synchronous, hybrid, and face-to-face teaching/learning. IJTE **4**(1), 46–56 (2021)
8. Prieto, L.P., et al.: Supporting orchestration of CSCL scenarios in web-based distributed learning environments. Comput. Educ. **73**, 9–25 (2014)
9. Wang, P., et al.: Chao: a framework for the development of orchestration technologies for technology-enhanced learning activities using tablets in classrooms. IJTEL **10**(1/2), 1–21 (2018)
10. Sammons, P., et al.: Exploring variations in teachers' work, lives and their effects on pupils: key findings and implications from a longitudinal mixed-method study. BERJ **33**(5), 681–701 (2007)
11. Kayi-Aydar, H.: Teacher agency, positioning, and English language learners: Voices of pre-service classroom teachers. TATE **45**, 94–103 (2015)
12. Manathunga, K., Hernández-Leo, D.: Has research on collaborative learning technologies addressed massiveness? A literature review. ET&S **18**(4), 357–370 (2015)

13. Hull, M.M., et al.: Validation of a survey to measure pre-service teachers' sense of agency. In: Journal of Physics: Conference Series, vol. 1929, no. 1, p. 012085. IOP Publishing (2021)
14. Hart, S.G., Staveland, L.E. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Advances in Psychology, vol. 52, pp. 139–183. North-Holland (1988)
15. Leppink, J., van den Heuvel, A.: The evolution of cognitive load theory and its application to medical education. Perspect. Med. Educ. **4**(3), 119–127 (2015). https://doi.org/10.1007/s40037-015-0192-x
16. Shaffer, D.W., et al.: A tutorial on epistemic network analysis: analyzing the structure of connections in cognitive, social, and interaction data. JLA **3**(3), 9–45 (2016)
17. Emerson, L., MacKay, B.: A comparison between paper-based and online learning in higher education. BJET **42**(5), 727–735 (2011)
18. Amarasinghe, I., et al.: Deconstructing orchestration load: comparing teacher support through mirroring and guiding. IJCSCL **16**(3), 307–338 (2021). https://doi.org/10.1007/s11412-021-09351-9
19. Csanadi, A., et al.: When coding-and-counting is not enough: using epistemic network analysis (ENA) to analyze verbal data in CSCL research. IJCSCL **13**(4), 419–438 (2018). https://doi.org/10.1007/s11412-018-9292-z
20. Zhang, L., et al.: Examining different types of collaborative learning in a complex computer-based environment: a cognitive load approach. CHB **27**(1), 94–98 (2011)

Implicit HCI for Geocollaborative Hyperstories Creation

Nelson Baloian[1](✉), Gustavo Zurita[2], José A. Pino[1], and Rodrigo Llull[1]

[1] Department of Computer Sciences, Universidad de Chile, Beaucheff 851, Santiago, Chile
{nbaloian,jpino}@dcc.uchile.cl

[2] Department of Information Systems and Management Control, Faculty of Economics and Business, Universidad de Chile, Diagonal Paraguay 257, Santiago, Chile
gzurita@fen.uchile.cl

Abstract. To create a geo-collaborative hyperhistory, physical areas associated with data and multimedia content are geolocalized over a map, from which links to other areas can be generated, which define paths of exploration of the hypernarrative. In this work in progress, we aim at facilitating the creation of geo-collaborative hyperstories, by redesigning the HCI of an existing application using implicit HCI principles. Implicit HCI (iHCI) advocates using the user's context information to anticipate the actions they want to perform, facilitating interaction and alleviating their cognitive load. iHCI has usually used as single-user interaction; therefore, we explore ways to extend its reach by taking contextual information from a group that works collaboratively. The result is a redesign proposal of six frequent tasks during creating and reading a hypernarrative described according to the existing literature on iHCI.

Keywords: Geocollaboration · Hyperstories · Implicit human-computer interface

1 Introduction

A hyperstory can be seen as a collection of multimedia objects being part of a story and organized as a graph, allowing the reader to traverse it in various ways [1]. Geocollaboration concerns the collaborative process of artifact creation in which information is strongly related to geographical places [2]. Geocollaborative hyperstories have been frequently used to develop documents about travel reports, quest narratives, and biographies [3].

The authors have previously developed an application allowing the construction of geocollaborative hyperstories with cultural heritage contents [4]. Users of this application are permitted to highlight an unlimited number of geographical sites with cultural interest on a map. Moreover, the users can connect these sites with arbitrarily labeled links. This feature allows users to relate areas according to thematic criteria and recommend possible visiting routes to other users. Each area (or site) can be enriched with multimedia material, text, links to external web pages, etc. The software has an interface

L.-H. Wong et al. (Eds.): CollabTech 2022, LNCS 13632, pp. 304–312, 2022.
https://doi.org/10.1007/978-3-031-20218-6_22

compatible with smartphones, so hyperstory builders can create hyperstory nodes on the map at the same real-world place of the cultural object, easing the tasks of knowing the exact geolocation and addition of multimedia material. The software was evaluated by six groups of 3 to 5 users; each group developed hyperstories on monuments, old churches, and museums, relating them according to hierarchy, location, recommended tour, or thematic criteria. A study on usability and utility showed that users evaluated the application as very useful; the task and tool increased their interest in cultural heritage. However, most participants in the experiment said creating these hyperstories was not an easy task since it required knowing in advance how to execute the actions for this purpose (e.g., adding multimedia material or relevant textual information obtained from the Internet), and these can be many, so this was an obstacle to develop high-quality documents.

This paper concerns a solution we propose to the problem mentioned above. We use the implicit interaction concept Field [5] to redesign the system; it is called iGeoHype (**I**mplicit HCI in **Geo**collaborative **Hype**rstories). We may note that "implicit interaction is about computers understanding the intentions and needs of the user and proactively triggering functions or adapting the interface to help users achieve their goals" [6]. This work in progress proposes actions that are implicitly activated when using iGeoHype, helping the user complete and improve her interaction with the application. The proposed redesign has not been evaluated yet. However, we believe the ideas behind the redesign could be helpful to other developers whose applications may have the same problem.

2 Related Work

2.1 Hyperstories in Culture Heritage Context

According to [5], a hyperstory is a powerful didactical tool since it can be used to present ideas and share knowledge through the integration and coherent organization of multimedia resources in various collaborative technological platforms [6]. Authors generate content by selecting a topic, researching it, creating a script, and developing a story [7]. Users can take advantage of collaborative spaces to jointly produce information pills that can be part of hyperstories, requiring a careful elaboration of a literary and technical script that integrates the options that hypertext allows. Relevant approaches to creating hyperstory processes are [7]: a) **linear and non-linear; they** differ based on the action sequences of the narrative that occur in the story. Non-linearity allows storytellers to tell complex stories with various stories within the same story; b) a **social/collaborative narrative** allows designing an active and participatory experience, which improves the narrative structure; and c) a **mobile/ubiquitous narrative** allows a narrative to take place in the physical environment where the story is generated or referenced, by using mobile devices and wireless connectivity. A context where hyperstories can be conveniently used is an appreciation of cultural heritage when creating stories around cultural sites, collecting data from historical data records, and combining it with informal knowledge transmitted by word of mouth through generations. In the field of education, research has been done on using hyperstories to support teaching/learning, e.g., in the Australian history [6], and to increase intercultural awareness in the higher education [8]. The description of cultural heritage objects may include videos, pictures (taken from the Internet or captured on-site), links to web pages with additional information, and georeferencing information.

Hyperstories construction can introduce the “getting lost in hyperspace” problem [9], which refers to the difficulty a user may face while building a common thread between a narrative with various links to other narratives. Our approach to this problem includes georeferencing on maps as a central scaffold and using implicit functionalities to ease human-computer interaction.

2.2 Geocollaboration and Hyperstories

Maps provide much information in a limited space. Today, a student, a teacher, a scientist, or anyone with access to a smartphone or computer can use a map to tell a story. Collaborative *storytelling* tools allow users to compile and show georeferenced data usually obtained from various sources. These tools are ArcGis StoryMaps or Knight Lab StoryMapJS, enabling users to create hyperstories and observe others that are open to the public. Following a hyperstory in the applications mentioned above makes a path through slides, each associated with a place in the hyperstory. There are mainly two approaches to navigating them: buttons and scroll. Only Knight Lab StoryMapJS creates arcs to represent hyperstory continuity through nodes. None of the applications use pop-up labels on the sites. No implicit interaction elements were found in the reviewed applications.

2.3 What is an Implicit Human-Computer Interface?

Traditionally, a user’s interaction with an application interface is explicit (eHCI), i.e., the user proactively tells the interface what to do through input mechanisms like direct manipulation, gestures, voice, etc. This requires a high engagement and participation from the user. On the contrary, implicit interaction (iHCI) means the application proactively generates actions, anticipating user needs [10]. iHCI captures and understands the user’s intentions by analyzing the context and her actions, resulting in an adaptable service within a dynamic context. According to [10], iHCI is based on the ability of the computer to understand the user requirements in a given context and its interpretation, responding in a meaningful and practical way without the user explicitly requesting it. Simple examples of use of iHCI principles are motion detectors that open and close automatic doors or the activation of automatic escalators when people approach them. More complex examples in education include using smart identification devices in classrooms to automate attendance records, seamlessly distribute material and learning activities to participating students, and ease interaction between teacher and students [11].

2.4 Implicit Human-Computer Interface in Storytelling

Storytelling has been used as an effective way to capture the motivation of formal and informal learners, improving their interest in exploring new ideas. On the other hand, hypermedia has been used to support learning through storytelling enabling users to create stories that can be read from various points of view [4].

Creating hyperstories is not a fast and easy task, and traditional tools like iGeoHype o Google Tour Builder do not include powerful tools to support this task. For instance, if a user needs to add information about the site during hyperstory creation, then she has to search for it on Internet; if she needs to check whether there is a similar hyperstory about the same site, then she must search for it within the application; if she wants to use some existing hyperstory as a basis for a new one, then she has to look for it too; If she does not have good pictures and would like to get one from Internet, then she has to do it outside of the application. As mentioned above, implicit interaction has been rarely used in contexts of storytelling with maps.

2.5 Design of iGeoHype to Support Hyperstories

iGeoHype is a web application for creating and visualizing georeferenced hyperstories in the context of cultural heritage appreciation through storytelling. Applications such as iGeoHype have been shown to be capable of increasing users' interest and knowledge of cultural heritage (iGeoHype, google-ToolBuilder). This application allows creating and visualizing hyperstories for users through a map, adding sites and links between them.

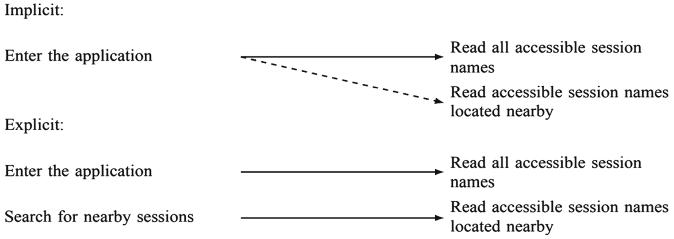
A **session** must be defined to begin the creation process. It is a map environment with the tools to create a hyperstory. Once inside, the user can create sites using an "add" button, adding different types of marks on the map (pins, circles, rectangles, user-defined areas) which conform to the nodes of the hyperstory. These elements or *sites* can be annotated with a title, description, images, and tags. The sites can be colored. Clicking on-site displays a popup menu where the information added at the time of its creation is displayed, as well as images and tags. These sites can be related to each other by links defined between them, thus building the hyperstory; see [4] for more details. An essential aspect of the application is geocollaboration, which allows various users to access the same session for viewing or editing. Users can also interact with other authors in a chat session. This work proposes to include implicit interaction in the storytelling of iGeoHype, aiming to improve navigation and the use of the platform to keep increasing cultural heritage interest and knowledge.

3 iHCI Proposal for Improving the Application

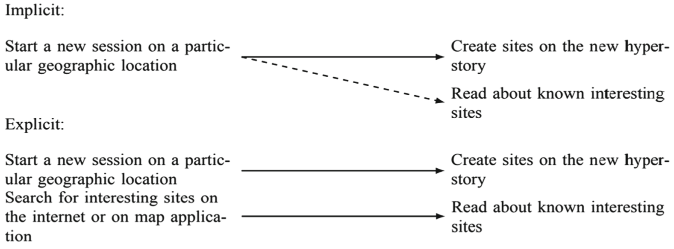
Serim & Jacucci propose and define a methodology for characterizing the types of iHCI interactions [5]. They present five types of implicit interactions: 1) **unintentional** actions triggered by the application to provide answers beyond what the user wanted; 2) **attentional background:** actions triggered by the system's detection of the user's surrounding environment; 3); **unawareness:** actions triggered by the system responding without the user being aware of them; 4) **unconscious:** implicit actions triggered without conscious processing of the user; and 5) **implicature**: interactions triggered by meanings that are implicit in the user's expression and communication. Below we propose, explain and justify concrete iHCI instances to be introduced to iGeoHype, as a redesign of the application presented in [4], based on the implicit interaction classification described in [10]:

i. Using the **geolocation** of the user or the sites of a hyperstory as *information*, we propose the following actions as implicit interactions:

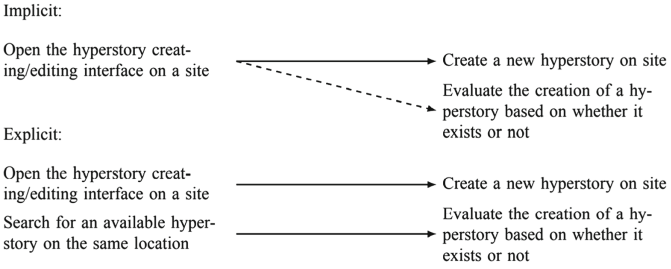
a) Suggesting existing sessions having sites (nodes of the hyperstory) georeferenced nearby. For instance, when the user searches for sessions, the application highlights the names of those hyperstories with elements georeferenced near the user's location on the list. Its characterization is: a) **Unintentional** because the system understands what the user wants but has not explicitly requested. b) **Background** allows the user to focus on the usage of the application. According to [5], the way to explain and compare the implicit interaction (iHCI) to the necessary explicit interaction (eHCI) to achieve the same goal would be:


Implicit:
Enter the application
Read all accessible session names
Read accessible session names located nearby
Explicit:
Enter the application
Read all accessible session names
Search for nearby sessions
Read accessible session names located nearby

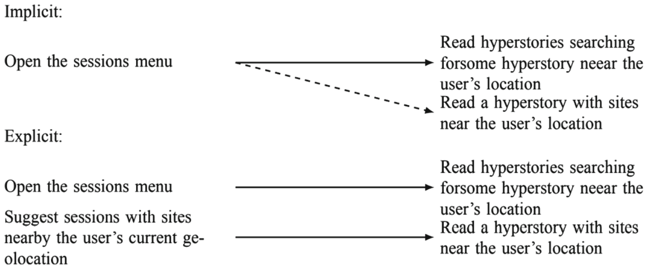
b) When creating/editing a hyperstory, the application may suggest sites of interest that are located nearby, not included in the hyperstory, and could be added as part of it. These sites may be obtained from the Internet. They may be classified as follows: a) **Unconscious**, because it reduces the user cognitive load of creating the hyperstory. b) **Implicature**, due to the same argument. c) **Background**, because it allows users to keep their attention on the creation task, thus achieving more productivity. Comparing the implicit vs. explicit interaction:


Implicit:
Start a new session on a particular geographic location
Create sites on the new hyperstory
Read about known interesting sites
Explicit:
Start a new session on a particular geographic location
Create sites on the new hyperstory
Search for interesting sites on the internet or on map application
Read about known interesting sites

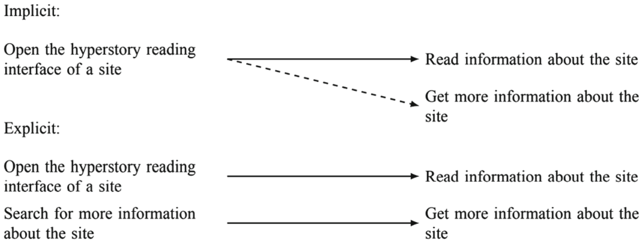
c) When adding a site to a hyperstory the application may inform the user the presence of another hyperstory nearby. The user could then reuse the information contained in the existing site for the hyperstory she is creating. Its characterization is: a) **Implicature**, because with this functionality the user may not need to create new material, re-using existing one. This reduces the necessary effort for creating a hyperstory. b) **Unintentional** because the user may not need it, but this can help. Explaining and comparing the explicit vs explicit interaction:


Implicit:
Open the hyperstory creating/editing interface on a site
Create a new hyperstory on site
Evaluate the creation of a hyperstory based on whether it exists or not
Explicit:
Open the hyperstory creating/editing interface on a site
Create a new hyperstory on site
Search for an available hyperstory on the same location
Evaluate the creation of a hyperstory based on whether it exists or not

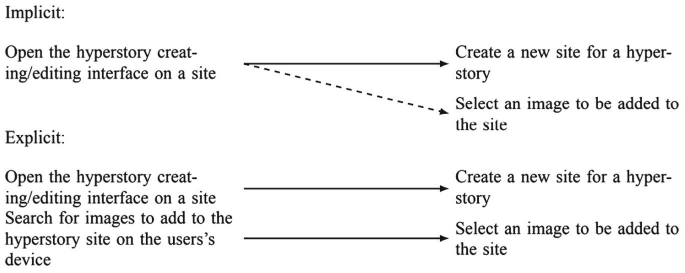
d) When exploring a site content, the application can suggest the users explore other hyperstories with locations nearby the one they are exploring. Its characterization is the following: a) **Unintentional**, because the application understands what the user needs without having to request it. b) **Background**, because it allows the user to focus on the task. Explaining and comparing the implicit vs. explicit interaction:


Implicit:
Open the sessions menu
Read hyperstories searching forsome hyperstory neear the user's location
Read a hyperstory with sites near the user's location
Explicit:
Open the sessions menu
Read hyperstories searching forsome hyperstory neear the user's location
Suggest sessions with sites nearby the user's current geolocation
Read a hyperstory with sites near the user's location

ii. Using the **name of a site** the user is exploring as information, the application could provide additional information about the site being read and downloaded from Internet. This may be useful when the user is interested in further information without being forced to change the application. Its characterization is the following: a) **Unintentional**, because the application provides the user with helpful information which has not explicitly been requested. b) **Background**, since it allows the user to focus on the site creation task. Explaining and comparing the implicit vs. explicit interaction:


Implicit:
Open the hyperstory reading interface of a site
Read information about the site
Get more information about the site
Explicit:
Open the hyperstory reading interface of a site
Read information about the site
Search for more information about the site
Get more information about the site

iii. Based on the **content** of a hyperstory site: While creating a place for a hyperstory, the application can suggest images, keywords for the title, or labels for the links among the sites. Its characterization would be: a) **Implicature** since it allows for the task of adding photos with less effort. b) **Unintentional** because it is not explicitly requested by the user but may help. Explaining and comparing the implicit vs. explicit interaction:


Implicit:
Open the hyperstory creating/editing interface on a site
Create a new site for a hyperstory
Select an image to be added to the site
Explicit:
Open the hyperstory creating/editing interface on a site
Search for images to add to the hyperstory site on the users's device
Create a new site for a hyperstory
Select an image to be added to the site

The proposals can be summarized as presented in Table 1.

Table 1. Summary of proposals

	Geolocation	Content
Reading	Suggest reading other hyperstories containing sites geolocated nearby or presenting internet pages about objects located nearby	Suggest reading other hyperstories containing sites with similar content (text, images, etc.) or presenting internet pages about objects with similar content
Creating	Suggest adding content form sites of other hyperstories located nearby or material from the internet about objects located nearby	Suggest adding content form sites of other hyperstories with similar content (text, images, etc.) or material from the internet about objects with similar content

Some examples of how the implicit interaction looks on the user's mobile phone interface are shown. We consider the cases when the application makes an Internet search for material related to a particular location. Figure 1. shows how a hyperstory is created with places and links among them [4]. Figure 2.. Shows how the application presents suggestions for keywords and descriptions for a particular place, based on the name of the area given by the user (here Saint Jaume). Description and keywords are retrieved from the Internet and sites created by other users. Figure 3. shows how the application suggests including images taken from the Internet and projects previously created by other users, which can be added to the site by simply pressing a button.

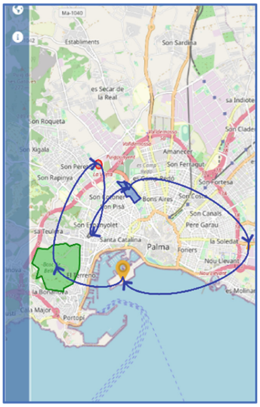
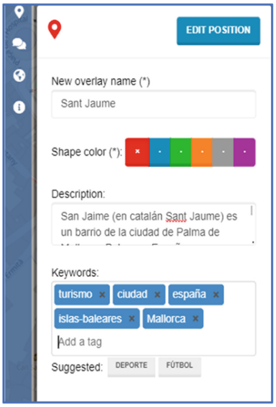


Fig. 1. Creation of georeferenced sites and links.



EDIT POSITION

New overlay name (*)

Sant Jaume

Shape color (*):

Description:

San Jaime (en catalán Sant Jaume) es un barrio de la ciudad de Palma de

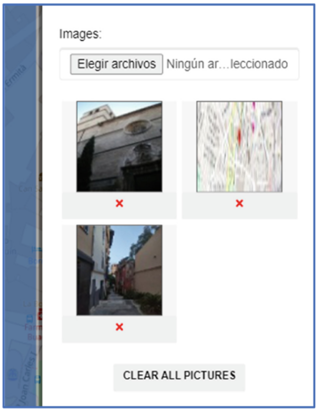
Keywords:

turismo × ciudad × españa × islas-baleares × Mallorca ×

Add a tag

Suggested: DEPORTE FÚTBOL

Fig. 2. Suggestion of keywords and description.



Images:

Elegir archivos Ningún ar...leccionado

CLEAR ALL PICTURES

Fig. 3. Suggestion of images.

4 Conclusions

The literature shows georeferenced hyperstories may be a powerful tool, with localization playing a relevant role. They can be used to support learning activities and informative presentations. Their users can share knowledge in the place where it is generated through coherent collaborative integration and organization of the georeferenced multimedia resources. Baloian et al. [4] presented the development of an application allowing the construction of collaborative hyperstories with cultural heritage contents; studies done with real users indicated that HCI design needed improvement. This paper concerned precisely this goal. We proposed using the implicit interaction concept (iHCI) [10] to improve the interaction between the user and the iGeoHype application. iHCI implies the system should detect user intentions and needs and then activate functions pro-actively, adapting the interface to help users in their activity.

This work proposes incorporating at least six concrete iHCI actions into the iGeoHype design. These actions are classified in three relevant instances when the user is interacting with the application: 1) using the user's **geolocation** or the sites of a hyperstory as information, 2) using the **name of a site** the user is exploring as information; the application could provide additional data about the site being downloaded from the Internet, and 3) based on the **content** of a hyperstory site. These instances correspond to basic actions. Other instances could be added.

References

1. Charles, V.M.: I'll Tell the Story My Way! Multi-perspective, Multimodal Storytelling in an Elementary Classroom (2018)
2. Câmara, J.H., Vegi, L.F., Pereira, R.O., Geöcze, Z.A., Lisboa-Filho, J., de Souza, W.D.: ClickOnMap: a platform for development of volunteered geographic information systems. In: 2017 12th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1–6. IEEE (2017)

3. Mayr, E., Windhager, F.: Once upon a spacetime: visual storytelling in cognitive and geotemporal information spaces. ISPRS Int. J. Geo Inf. **7**, 96 (2018)
4. Baloian, N., Zurita, G., Pino, J.A., Peñafiel, S., Luther, W.: Developing hyper-stories in the context of cultural heritage appreciation. In: Nakanishi, H., Egi, H., Chounta, I.A., Takada, H., Ichimura, S., Hoppe, U. (eds.) Collaboration Technologies and Social Computing: 25th International Conference, CRIWG+CollabTech 2019, Kyoto, Japan, September 4–6, 2019, Proceedings, pp. 110–128. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-28011-6_8
5. Robin, B.R.: Digital storytelling: a powerful technology tool for the 21st century classroom. Theory into practice **47**, 220–228 (2008)
6. Smeda, N., Dakich, E., Sharda, N.: Digital storytelling with Web 2.0 tools for collaborative learning. In: Collaborative Learning 2.0: Open Educational Resources, pp. 145–163. IGI Global (2012)
7. Gaeta, M., Loia, V., Mangione, G.R., Orciuoli, F., Ritrovato, P., Salerno, S.: A methodology and an authoring tool for creating complex learning objects to support interactive storytelling. Comput. Hum. Behav. **31**, 620–637 (2014)
8. PM Ribeiro, S.: Developing intercultural awareness using digital storytelling. Lang. Intercultural Commun. **16**, 69–82 (2016)
9. McLellan, H.: Hyper stories: some guidelines for instructional designers. J. Res. Comput. Educ. **25**, 28–49 (1992)
10. Serim, B., Jacucci, G.: Explicating “Implicit Interaction” an examination of the concept and challenges for research. In: Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems, pp. 1–16 (Year)
11. Alvarez, C., Zurita, G., Baloian, N.: Applying the concept of implicit HCI to a groupware environment for teaching ethics. Pers. Ubiquit. Comput. (2021). https://doi.org/10.1007/s00779-020-01495-z

Author Index