Journal of Geography in Higher Education

ISSN: 0309-8265 (Print) 1466-1845 (Online) Journal homepage: www.tandfonline.com/journals/cjgh20

A comparative study on the use of spatiotemporal maps for teaching data visualization in a hybrid learning context

Cristina Portalés, Javier Sevilla, Pablo Casanova-Salas, Mar Gaitán & Sergio Casas

To cite this article: Cristina Portalés, Javier Sevilla, Pablo Casanova-Salas, Mar Gaitán & Sergio Casas (2025) A comparative study on the use of spatiotemporal maps for teaching data visualization in a hybrid learning context, Journal of Geography in Higher Education, 49:3, 332-355, DOI: 10.1080/03098265.2024.2423408

To link to this article: https://doi.org/10.1080/03098265.2024.2423408

JOURNAL OF GEOGRAPHY IN HIGHER EDUCATION
2025, VOL. 49, NO. 3, 332–355
https://doi.org/10.1080/03098265.2024.2423408

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

A comparative study on the use of spatiotemporal maps for teaching data visualization in a hybrid learning context

Cristina Portalés[a], Javier Sevilla[a], Pablo Casanova-Salas[a], Mar Gaitán[b] and Sergio Casas[a]

[a]Institute of Robotics and Information and Communication Technologies, Universitat de València, Valencia, Spain; [b]Department of Art History, Universitat de València, Valencia, Spain

ABSTRACT
Data visualization is a complex matter and so is teaching it in the context of engineering and data science. Even more challenging is teaching in a midst of once in a century pandemic. In this paper, we show the results of a workshop conducted with second-year students on a university degree in data science, which took place in the context of pandemic restrictions in a hybrid learning environment. About half of the students were in the classroom and the other half at home. They learned to use STMaps, a tool developed in the scope of the SILKNOW project. We transferred research outcomes to the classroom; similarly, the results of evaluating the tool were transferred to the project. But more importantly, we found insights that the learning process of students that were online was not the same quality as of those that were physically in the classroom. While both groups performed well in terms of correctly fulfilling the tasks proposed during the workshop, the students at home were less motivated and perceived the tasks more cumbersome to achieve. From a research perspective, this activity was perceived as very useful, and similar activities are being planned for each academic year.

ARTICLE HISTORY
Received 8 November 2023
Accepted 26 October 2024

KEYWORDS
Data science; online learning; data visualization; knowledge transfer; spatiotemporal maps

Introduction

The boundaries between teaching and research have been crossed several times within the academic environment. While research involves the creation of knowledge, for example the advancement of a technology with respect to the state of the art, teaching involves the transfer of knowledge, usually to young generations that will be the takers of such knowledge. Both activities might benefit from each other if they are well aligned.

In the scope of the research project SILKNOW (Portalés et al., 2018), we have developed STMaps (Spatio-Temporal Maps, Figure 1), a tool for the spatiotemporal visualization of data related to silk heritage and their relationships, using a combination of map-based tools that retrieve information from a knowledge graph. The details of the design and implementation of this tool were previously reported in

CONTACT Cristina Portalés Cristina.Portales@uv.es Institute of Robotics and Information and Communication Technologies, Universitat de València, Valencia, Spain

Supplemental data for this article can be accessed online at https://doi.org/10.1080/03098265.2024.2423408

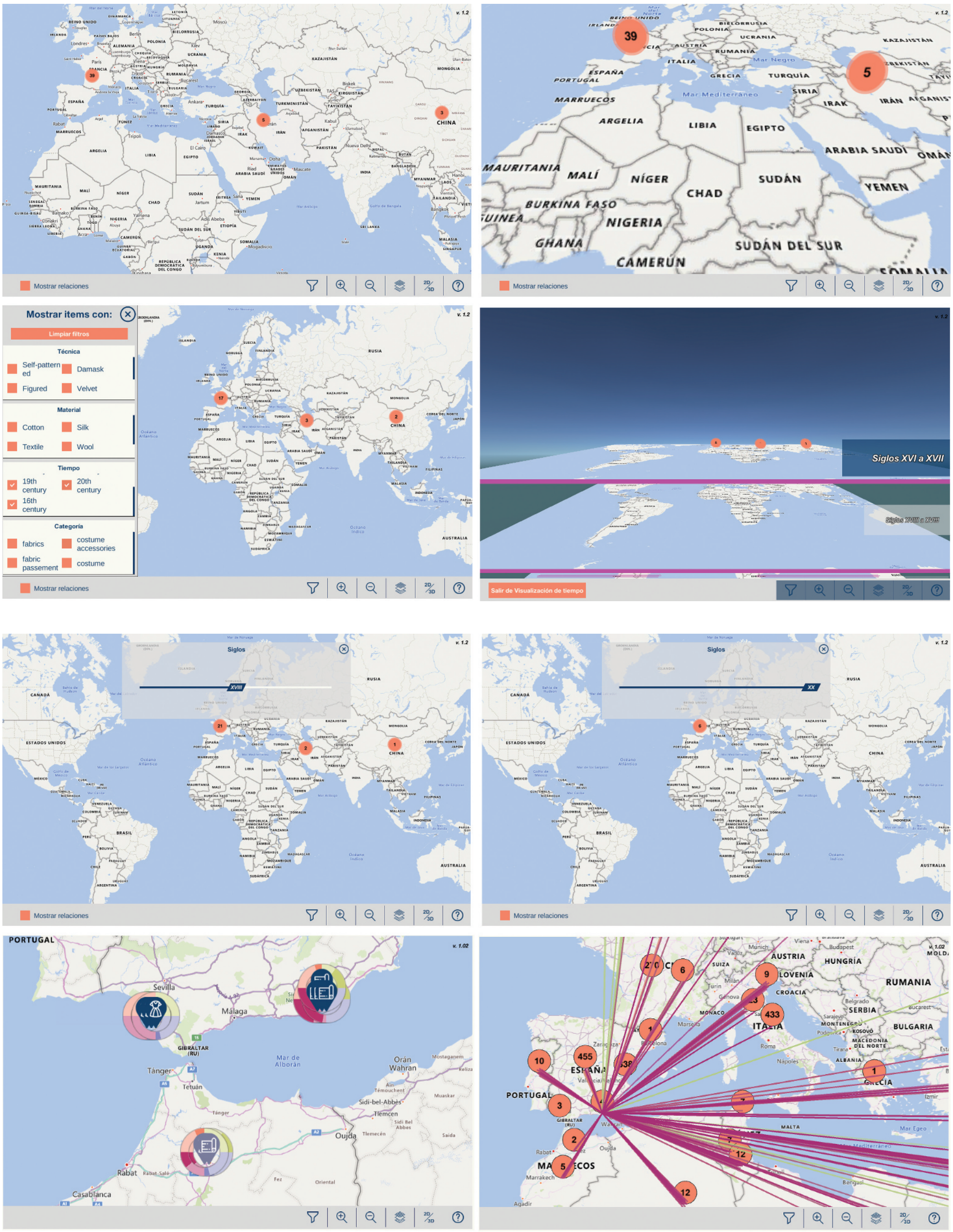

Figure 1. Examples showing different visual aspects of STMaps, where: (a) 2D markers; (b) 3D markers; (c) Temporal filtering; (d) Layered timescale; (e) and (f) Timeline display; (g) Ring-based relationships; (h) Line-based relationships.

(Sevilla et al., 2021). In contrast, in this study, we introduce the outcomes of integrating STMaps in the classroom. To that end, the tool was shown to and evaluated by second-year university students in a Data Science undergraduate program as part of a subject on data visualization. In this subject, students learn how to visualize data, not only with charts but also with interactive maps. Specifically, they learn about spatial data types, spatial (basic) analysis, and visualization with a Geographical Information System (GIS), and spatial visualization and interactive maps in R. Every year, as part of the subject plan, it is included a workshop session in which students are introduced to a real-world case, usually within the context of research. The goal is to present them with real-life scenarios that go beyond the technologies covered in class, providing them with insight into the professional world; this allows students to see the practical application of the concepts addressed in the course.

Additionally, by engaging students in the evaluation process of the SILKNOW project, we aim to strengthen the connection between research and teaching. Through this approach, students gain the opportunity to serve as one of the initial audiences to interact with the developed application and benefit from the knowledge generated. Researchers provide detailed explanations of the application's construction, thereby enhancing students' understanding of data visualization. Moreover, student feedback is valuable to researchers, as it can inform improvements to the application and guide future developments.

Therefore, in one of the sessions, a dedicated workshop was organized where the project and the developed tool were explained to the students. Then, they performed specific tasks with the tool and completed a questionnaire that included tests of perceived workload and system usability. The students' feedback was relevant to improving the tool during project life, thus benefiting the research activity. On the other hand, students benefited from the experience of having access to a technology under development, which included novel methodologies for visualizing data and their semantic relationships in a way beyond the available software. Therefore, this was a win-win activity, mutually benefiting research and the learning process.

The workshop was conducted during the COVID-19 pandemic. At that time, government restrictions imposed a limit on the number of students that could physically attend a classroom simultaneously, whereas the number of enrolled students was not reduced. This led us to develop classes in a hybrid mode with blended learning, so that some of the students were at home and some of the students were in the classroom. In this context, the workshop was conducted in a hybrid mode (face-to-face vs. remote). About half of the students were physically present in the classroom ($n = 19$, group A), while the rest were at home ($n = 15$, group B) using online teaching resources.

The aim of this research was threefold: First, to transfer research outcomes to the classroom (aim 1); Second, to transfer back students' feedback to the research activity (aim 2); and last, but not least, to assess the quality of the learning process in a hybrid classroom (aim 3). Related to aim 3, we wondered if, in such a hybrid context, the learning process of students that were online could be of the same quality as of those that were physically in the classroom.

This study focuses on describing the workshop and analyzing the results, and also compares the differences between the two groups for five parameters: (1) correctness, (2) dedicated effort or motivation, (3) visual preferences, (4) perceived usability, and (5) perceived effort. The analysis includes both qualitative and quantitative analyses, which are aligned with the three research aims and provide insights into the quality of the learning process in such a hybrid context.

The remainder of this paper is organized as follows. In section 2 we bring the related work, focusing on works that compare online vs. face-to-face learning, and explaining why our work is different from other studies. Section 3 is dedicated to the material and methods, where we explain the workshop in detail and the methodology to process the data. Then, section 4 analyses the results of the considered parameters, also comparing the responses of the two groups of students. Finally, we bring a discussion and a conclusions section.

Related work

The use of technology in the classroom is not new and has been discussed by a great number of authors in the academic literature (Bower, 2019; Conelly & Miller, 2018; Dunn & Kennedy, 2019; Kemp et al., 2019). Already before the COVID-19 outbreak, educational trends in higher education were to implement digital tools in the classrooms (EDUCAUSE, 2018) as some authors have addressed. Special attention has been given to those models where a significant part of teaching is online and is designed from its inception. We must outline this fact, as we will mention later, the COVID-19 crisis prompted a series of online and other hybrid models that respond more to emergence models than to pre-conceived pedagogical models. We are talking about blended learning (which combines both digital and face-to-face teaching) and flipped learning (which allows students to have access to lesson content before school, traditionally it has been considered the next step of blended learning). Both have been discussed and evaluated with positive attitudes towards them (Bond, 2020; Capone et al., 2017; Halasa et al., 2020; Tang et al., 2020; Thai et al., 2017; Vallée et al., 2020).

On the other hand, other research has compared online learning with face-to-face courses. These studies have dealt mostly with academic performance, student engagement and student satisfaction. Results vary according to course type, characteristics of the target group (demographics, socio-economic data), type of tools used, etc. In this regard we can mention some studies such as the ones performed by Elfaki et al. (2019) who compared both online learning with face-to-face on a nursing degree and whose results showed that those e-learners had higher performance than those who had traditional face-to-face teaching. Bailey (2020) compared face-to-face, hybrid and online modalities at a US public university demonstrating that there are no significant differences in these modalities. Soffer and Nachmias (2018) at a university in Tel Aviv, compared the effectiveness of three online courses with the same face-to-face courses. Online students reported a better understanding of the course structure and higher engagement and satisfaction. However, face-to-face students reported that there was a better contribution of the course to their learning. Daniela et al. (2018) conducted research to evaluate the extent of technology in education; their results showed a correlation between the use of Information and Communication Technologies (ICT) in students' engagement and their active learning. However, they also recognize that it depends on student's age and their digital literacy. In this regard, Bond et al. (2018) state that using technology solely does not act as a driver of learning as face-to-face can be equally as or more effective. They performed a survey in a German University where they found that students are more likely to embrace digital tools, but its use depends much on the teachers' digital literacy and willingness to adopt them.

Despite the above-mentioned results, it is important to mention that these studies were mostly conducted before the pandemic hit the world. As we all know, the COVID-

19 outbreak has disrupted higher education from on-site education to exclusively online models. Protocols around the world included severe lockdowns or at least some regulations to contain the outbreak. This situation forced education to become remote and activate what Bozkurt and Sharma (2020) have called Emergency Remote Teaching (ERT). This means that online education was set up as a response rather than a permanent solution. In fact, while the virus was contained and people got vaccinated, some Universities decided to try hybrid systems, that is having some students at home and others in the classrooms. Translating it to reality meant that people (teachers and students included) had limited time for responding and adapting their learning (Patricia Aguilera-Hermida, 2020).

This type of online technology applied to teaching and learning has also been explored from several angles. For example, the perceived usability of the tools used by Pal and Vanijja (2020), who demonstrate that the Technology Acceptable Model (TAM) and System Usability Scale (SUS) (Brooke, 1996) scale are correlated when analyzing Microsoft Teams as a learning platform during the pandemic; however, they did not perform any analysis related to learning or how the sustainability of the use of it backed to face-to-face teaching. Other studies focused on academic performance with different results. Some had positive responses such as the one conducted by Alvarez-Herrero and Hernández Ortega (2020) who compared the academic performance of online courses with the same face-to-face education degree courses, being the ones online significantly better than those who were face-to-face. Nevertheless, the study compares the course 2018–2019 (mostly pre pandemic) and 2019–2020 (during the pandemic). Indeed, they recognize that still is too early to assess positively these results, especially considering that the pandemic has had impact in everyone's lives. Others encompassed a wider range of aspects like the one done by Faura-Martínez et al. (2021) who performed a study on the impressions of Spanish university students during the Pandemic. Among their findings, we raise the fact that students found online classes and activities through the virtual platform as one of the most useful resources. In this study, they also point out that most of the students (53%) found that their academic performance has been negatively affected by online teaching.

When examining studies that assess the utility of Geographic Information Systems (GIS) technologies in hybrid and e-learning educational approaches, it is commonly found that students achieve better academic results in hybrid education compared to e-learning (Safwa et al., 2022; Zhang et al., 2022). Studies on the use of GIS technologies in teacher training (Moore et al., 2016) reveal a strong correlation between the implementation of these technologies and the self-reported preparedness of educators. However, other research (Kurniawan et al., 2022), questions these findings and suggests that effectiveness has not been adequately assessed considering factors such as context, knowledge, or skills, and that reassessment yields contrary results. Therefore, further research and analysis are required to confirm that the use of GIS technologies with a hybrid educational approach is more advantageous than using them with other educational approaches.

As it can be noted, these studies compare online teaching with face-to-face courses, as results vary and depend on several factors (age, digital literacy, socioeconomic structure, etc.) it might be too early to establish conclusions. On the other hand, we must point out that they compare full courses rather than specific activities such as the one we are dealing with. Nonetheless, as Xing and Marwala (2017) suggest, with the Fourth Industrial Revolution arising, digital learning is becoming more and more a reality as it facilitates critical thinking (especially through blended learning) and is innovative, facilitating research from students towards researchers and vice versa through exploring new research domains. In this sense, according to Jenkins (2000) an effective way to link research and teaching is to design courses where students benefit from staff research. Like this, knowledge transfer can be ensured and be transformed into evidence that solves a problem or improves a theory or product (Verhagen et al., 2014). Our paper, besides comparing online vs face-to-face results in a teaching activity during the pandemic, demonstrates how thanks to these types of evaluation, technological tools developed by the researchers were improved.

Material and methods

Participants

The total number of students involved in the subject of data visualization was 42, who were randomly assigned to two groups of (originally) equal size: group A (to be physically present in the classroom) and group B (remote, working from home). Of these, 34 students showed up for the workshop (thus, $N = 34$) and gave their consent, 19 (55.88%) belonging to group A and 15 (44.12%) to group B. They were aged between 19 and 21 years old; 24 were male (11 in group A, 13 in group B), and 10 were female (8 in group A, 2 in group B).

Procedure

The workshop was organized during a two-hour session of the subject Data Visualization at the University of Valencia (Spain), that is integrated in the second year of the Bachelor's degree in Data Science. The workshop consisted of three parts: (1) explanation of the project (30 min); (2) explanation of tasks, the document they had to fulfil, and the purpose of the evaluation (15 min); and (3) realization of the evaluation (max. of 75 min).

During the first part, that took the first 30 minutes of the workshop, the research project was presented to all the students, i.e. to group A and group B simultaneously. This was carried out by the teacher of the subject, which additionally is one of the principal investigators of the project. She prepared a Power Point presentation of the project that group A could see on the projection screen, while group B could see the same information using Microsoft Teams, as the teacher shared her laptop's screen. While explaining the project, she put special emphasis in the STMaps tools and the visual strategies that were implemented to visualize the data, considering their spatiotemporal dimensions, as well as to visualize their relationships. Given the nature of the course, it was of relevance for the students (Figure 1): the possibility to visualize data in 2D vs. 3D; the different strategies to deal with the temporal variable, namely, to apply a temporal filtering, to show data arranged in stacked century-maps (layered timescale), or to navigate through centuries with a slider (timeline display); and the different strategies to visualize relationships, namely, ring-based and line-based.

During the second part, that took around 15 minutes, the tasks that the students had to realize, the document that they had to fulfil, and the purpose of the evaluation were explained, also by the teacher and following the same procedure as in the first part. Students were informed that they had to fulfil some tasks using STMaps, which were described in a document, and then respond to the questions related to each task. They were also informed that, after completing the tasks and questions, they had to complete two questionnaires, one for the usability testing of the tool, and another to evaluate their perceived efforts. Such questionnaires were at the end of the document with the tasks and questions, and the meaning of the questions was explained. Regarding a consent form, they were informed that the fulfilled documents (that were anonymized), could be used by the people involved in the project with research purposes, but they could decide not to take part without any penalization in relation to the subject score. In this way, they were informed that all participants in the workshop would obtain the score associated to it (that represented 1% of the subject global score), regardless that they refused said researchers analyse their data. None of the participants refused that their data could be used with research purposes, neither during the workshop session nor with written notice to the teacher.

During the third part of the workshop, that took a maximum of 75 minutes, they had to do the tasks and complete the document with their answers to the tasks and questionnaires. The documentation is provided in the appendices: Appendix A (tasks and questions), Appendix B (questionnaire: NASA TLX Scale) and Appendix C (questionnaire: System Usability Scale). Note that for the purpose of the current paper, the texts have been translated from Spanish to English.

Both groups had the same access through a Moodle e-learning platform. They had to perform a total of 10 tasks and respond between 1 and 4 questions per task. During this third part, students were allowed to do short breaks if required. Students that finished before the time was over, could leave the classroom. In addition to the teacher, there were two other persons from the technical staff of the project, who were the responsible for the implementation of STMaps tool. One of them was physically in the classroom to assist students of group A in case they experienced any technical problem, as the STMaps tool was under development (although in a very final phase). Similarly, the other technician was at his office, connected via Microsoft Teams to assist students in group B.

Apparatus

Students in both groups completed the tasks and fulfilled the related document with their own laptops; the authors assume this to be the same hardware that they use for the rest of the subjects of the degree. This means, although there can be differences between the internal capabilities of the students' machines and their operating system, they should be sufficient to accomplish the tasks. Additionally, the students that were at home (group B) were connected to the classroom via videoconference, making use of the software Microsoft Teams, which is the tool provided by the University of Valencia for remote teaching and that was widely used by all the students during the pandemic. Therefore, students were already familiar with this tool.

Data analysis

We performed the experiment using a between-subjects design with two treatments. The independent variable is the learning method (face-to-face vs. online, i.e. group A vs. group B) and the dependent variables, the data that we measure in the questionnaires. The reason for using this method is that a between-subjects design minimizes the learning and transfer across conditions and requires a shorter time, which is important given the short amount of time we had for this workshop.

From the collected data, we were able to measure five quantitative parameters, namely, (1) the correctness, (2) the dedicated effort or motivation, (3) the visual preferences, (4) the perceived usability, and the (5) perceived effort. In Table 1, a list of mappings of measured parameters in relation to the questions and aims of this work are depicted; the related questions to each parameter are described in the following lines, whereas their relation to the aims of this work is justified in the discussion section. Additionally, we did a qualitative analysis, which is also summarized in Table 1. All these parameters were derived for the two groups.

The correctness was calculated from the answers given to questions Q7.2, Q7.3, Q7.4, Q8.1, Q8.2, Q8.3 and Q9.1 (see Appendix A), because the answers to these questions can be objectively correct or incorrect. From the answers given by the students to each question, a score of one was assigned if the answer was correct, zero otherwise. Then, the sum of correct answers for each question and group was calculated.

The motivation, or dedicated effort, was assessed using different strategies and making use of different sets of questions (see Appendix A):

- The questions Q3.1.1, Q9.2 and Q10.1.1, which required that students gave a motivated answer and that were mandatory. To rate the effort, the teacher assigned a value of 1 if she considered that the answer given by the student was motivated, and zero otherwise. For instance, examples that the teacher considered motivated answers are (translated from Spanish):

> Q3.1.1: The 3D visualization, I think, does not fully inform what you are trying to explain. Also, when you move around the map the markers crash and numbers disappear, or lines appear.
>
> Q9.2: I do not correctly understand the pattern in its distribution, since I have clicked on one of the related objects and no information appears around the material of said object, and therefore I do not know if said objects have been related to the former object for its material.

Table 1. Measured parameters in relation to the question and research aims.

Measured Parameters	Related Questions	Related Aim
Correctness	Q7.2, Q7.3, Q7.4, Q8.1, Q8.2, Q8.3, Q9.1	Aim 1, Aim 3
Motivation	Q3.1.1, Q9.2, Q10.1.1, Q7.1	Aim 3
Visual aspects	Q3.1, Q10.1	Aim 1, Aim 2
Perceived usability	SUS (all questions)	Aim 2, Aim 3
Perceived effort	TLXS (all questions)	Aim 2, Aim 3
Qualitative analysis	Q3.1.1, Q10.1.1	Aim 3

Q10.1.1: I think it is convenient to use filtering because the markers between the different objects are displayed and differentiated.

Whereas examples of non-motivated answers are (translated from Spanish):

Q3.1.1: The map is best seen in 2D.

Q9.2: It is concentrated more in Spain, France and Portugal.

Q10.1.1: It is displayed better.

- The average time dedicated to accomplishing Task 7 for each of the groups, which correspond to question Q7.1. Said time could be indicative of two facts: students find the task more difficult to accomplish it correctly, so they had to invest more time; or students were more motivated and thus (voluntarily) dedicated more time. However, as the students did not receive any reward for correctly accomplishing the tasks and given the fact that the sooner they finished, the sooner they could leave home, this study considers that the more time a student invested to complete Task 7, the more motivated he/she was.

The visual aspects were measured according to questions Q3.1 and Q10.1: Question Q3.1 was related to the preferred dimension (2D vs. 3D), where students had to select only one of the two options; Q10.1 was related to the preferred time selection (temporal filtering, layered timescale, and timeline display), where they could select more than one option.

For assessing the perceived usability when using the application, the answers given to the SUS were analysed using standard procedures. On the other hand, for assessing the perceived efforts, the answers given to the TLXScale were considered as well. Both SUS and TLX questions were asked for the task set as a whole, not for individual tasks.

To qualitatively assessing visual aspects, we analysed questions Q3.1.1 and Q10.1.1, where students had to develop in their own words, why did they choose one or another visualization. We grouped the answers in a total of 5 codes for Q3.1.1 and 7 codes for Q10.1.1. Grouping was done following an inductive method. That is, coding student's answers according to similar content (Kyngäs, 2020), every code is important as the other (He & Schonlau, 2020). Afterwards, we studied both questions solely and comparing both groups.

Results

We present the results of the study regarding the five quantitative parameters (correctness, dedicated effort or motivation, visual preferences, perceived usability, and perceived effort) and the qualitative analysis for two group of students: group A (face-to-face, n = 19, 55.88%) and group B (remote, n = 15, 44.12%). Previously, we performed a normality test as shown in Table 2, for those parameters that are objectively comparable. Shapiro-Wilk and Kolmogorov-Smirnov tests were chosen. They are the most widely used statistical tests for checking normality. In this case, the Shapiro-Wilk is considered more appropriate, as it works well for small sample sizes (<50 samples). Results show that the dataset measuring the time of Task 7 for group A fails to pass the Shapiro-Wilk

Table 2. Normality tests.

	Normality Test			
	Kolmogorov-Smirnov		Shapiro-Wilk	
Dataset	Statistic (D)	Significance	Statistic (W)	Significance
TLX1 – Group A	.148	.200***	.940	.260***
TLX1 – Group B	.141	.200***	.959	.681***
TLX2 – Group A	.206	.033*	.826	.003
TLX2 – Group B	.275	.003	.827	.008
TLX3 – Group A	.177	.120***	.967	.706***
TLX3 – Group B	.104	.200***	.939	.372***
TLX4 – Group A	.188	.075**	.857	.009
TLX4 – Group B	.195	.130***	.920	.190***
TLX5 – Group A	.227	.011*	.763	.000
TLX5 – Group B	.153	.200***	.955	.602***
TLX6 – Group A	.136	.200***	.911	.079**
TLX6 – Group B	.214	.062**	.884	.055**
TLX-Avg – Group A	.115	.200***	.919	.110***
TLX-Avg – Group B	.135	.200***	.927	.248***
SUS Score – Group A	.250	.003	.870	.015*
SUS Score – Group B	.120	.200***	.977	.948***
Time Task 7 – Group A	.169	.157***	.800	.001
Time Task 7 – Group B	.138	.200***	.950	.527***

Degrees of freedom = 19 (group A), 15 (group B).

test, although it passes the Kolmogorov-Smirnov test. Some of the datasets coming from the NASA-TLX (Hart & Staveland, 1988) also fail to pass this test, but the average NASA-TLX does pass it. Only the dataset for NASA-TLX 2 of Group B fails to pass both normality tests.

For those datasets that passed both normality tests, we applied parametric tests, such as the unpaired t-test, to compare the results from both groups. For those datasets that did not, we applied also non-parametric tests, such as the Mann-Whitney U test (also known as Wilcoxon rank-sum test) for independent samples. In addition, Fisher exact tests and Chi Square tests were applied to compare the observed and expected frequencies of answers. Significant results are divided into three categories: *** for $p \leq 0.01$, ** for $p \leq 0.05$ and * for $p \leq 0.1$ (except for the normality test in which the categories are inverted because the ≤ operator is changed by the ≥ operator).

Correctness

In this section, we assess if a given task was correct or not by analyzing the answer to questions Q7.2, Q7.3, Q7.4, Q8.1, Q8.2, Q8.3 and Q9.1, rather than judging student's attitudes towards it (Bandura, 1995).

As explained above, students were asked to access the map with their own laptop, and then to follow the tasks described in the document (Appendix A). After each task, they had to answer some questions. From this, the success rate, i.e. the percentage of correct answers, was computed, which is shown in Table 3. For group A, all questions obtain a success rate above 50%, and none reaches 100%. For group B, Q8.2 is below 50% whereas two questions reach 100%. The average success rate for group A is 83.46% (sd = 12.69%), and for group B 80.00% (sd = 23.41%). There are not significant differences between the two groups as seen in Table 4, where both Fisher and Chi Square

Table 3. Summary of correct answers for Q7.2-4, Q8.1-3 and Q9.1.

Questions	Group A (n)	Group B (n)	All (n)
Q7.2	**78.95%** (15)	73.33% (11)	76.47% (26)
Q7.3	84.21% (16)	**100.00%** (15)	91.18% (31)
Q7.4	**94.74%** (18)	93.33% (14)	94.12% (32)
Q8.1	94.74% (18)	**100.00%** (15)	97.06% (33)
Q8.2	**57.89%** (11)	33.33% (5)	47.06% (16)
Q8.3	84.21% (16)	**86.67%** (13)	85.29% (29)
Q9.1	**89.47%** (17)	73.33% (11)	82.35% (28)
Avg (sd)	83.46% (12.69%)	80.00% (23.41%)	81.93% (16.93%)

Table 4. Fisher and Chi Square tests for correctness.

Correctness Questions' Data						Test		
						Fisher	Chi Square	
Group	Correct	Correct (%)	Incorrect	Incorrect (%)	Total	Significance	Statistic χ^2	Significance
A	111	83.46%	22	16.54%	133	.5024	.4742	.4911
B	84	80.00%	21	20.00%	105			
Total	195	81.93%	43	18.06%	238			

tests reveal that the differences between the percentages of successful answers are not significant, so we can conclude that their performance was similar.

As the results show, with average success rates above 80%, students correctly understood how to use this tool. With these results, following (Davies, 2011) methodological framework for understanding and assessing technology literacy by providing three levels for understanding technology literacy (1: awareness; 2: praxis; 3: phronesis), we can say that the students were at the praxis level, where they are capable users who were guided in using our tool.

Dedicated effort/motivation

With respect to the motivation of the students, on the one hand Table 5 shows the summary of the motivated answers given to questions Q3.1.1, Q9.2 and Q10.1.1 in terms

Table 5. Summary of motivated answers for Q3.1.1, Q9.2 and Q10.1.1.

Questions	Group A (n)	Group B (n)	All (n)
Q3.1.1	**57.89%** (11)	33.33% (5)	47.06% (16)
Q9.2	**57.89%** (11)	33.33% (5)	47.06% (16)
Q10.1.1	**57.89%** (11)	40.00% (6)	50.00% (17)
Avg (sd)	57.89% (0.00%)	35.56% (3.85%)	48.04% (1.70%)

Table 6. Fisher and Chi Square tests for motivation.

Correctness Questions' Data						Test		
						Fisher	Chi Square	
Group	Motivated	Motivated (%)	Unmotivated	Unmotivated (%)	Total	Significance	Statistic χ^2	Significance
A	33	57.89%	24	42.11%	57	.0295**	5.0275	.0249**
B	16	35.56%	29	64.44%	45			
Total	49	48.04%	53	51.96%	102			

Table 7. Unpaired t-test and Mann-Whitney U tests for average times to complete task 7.

Test	Group	Mean ± SD	F (sig.)	T (sig.)
Unpaired t-test	A	46.89 ± 29.113	7.012 (.012)**	3.729 (.001)***
	B	18.20 ± 6.678		
Test	Group	Median ± IQR	U (sig.)	
Mann-Whitney U	A	40.0 ± 35.0	23.5 ($<10^{-3}$)***	
	B	19.0 ± 9.0		

Degrees of freedom = 32 (t-test).

of percentages for both groups and all the students. For each of the questions, group A gave more motivated answers. Both the Fisher and the Chi Square tests, shown in Table 6, reveal that the difference (57.89% vs 35.56%) between group A and B is statistically significant (p-value smaller than 0.05 for both tests).

On the other hand, the average time dedicated to accomplishing Task 7 for each of the groups, which corresponds to question Q7.1, was 46.89 seconds (sd = 29.11) for group A, and 18.20 seconds (sd = 6.68) for group B. Table 7 shows an unpaired t-test and a Mann-Whitney U test comparing group A and group B for this variable. Both tests reveal that the difference between the groups is statistically significant (p-value smaller than 0.01 for both tests). This reinforces the idea that students from group A were clearly more motivated to carry out the tasks they were prompted to do.

Visual aspects

Visual aspects were analyzed in terms of both the spatial and the temporal dimension. Regarding the spatial dimension, the map offers the possibility to view the base cartography and the markers in either 2D or 3D (see Figure 1(a,b)). After performing Task 3, students were asked to choose between the 2D and the 3D representation (Q3.1), where they could only select one of the two options.

On the other hand, the different strategies to deal with the temporal variable included to apply a temporal filter by selecting specific centuries on a side menu, as shown in Figure 1(c) (temporal filtering); to show data arranged in stacked century-maps, as shown in Figure 1(d) (layered timescale); or to navigate through centuries with a slider, with two examples depicted in Figures 1(e,f) (timeline display). All these options involve the addition and/or modification of graphical aspects related to the interface (e.g. the addition of a slide bar). After performing Task 10, the students were asked to select their preferred way to trigger the visualization of objects based on their temporal dimension, where multiple choices were possible (Q10.1).

The answers to Q3.1 and Q10.1 are shown in Figure 2 (note that the selected colour schemes for all the graphs in the paper are colourblind save). Regarding the spatial dimension, students in both groups preferred the 2D option; in fact, only one student in group A selected the 3D option. For the temporal dimension, there is a clear winner, temporal filtering, as 65% ($n = 22$) of all the students selected it, and it is also the most frequently selected option in each group individually. The second-preferred option is the timeline display, and the last-preferred option is the layered timeline. Comparing temporal filtering with the other two, it must be highlighted that this option allows the selection of a number of centuries that are not necessarily consecutive. This is

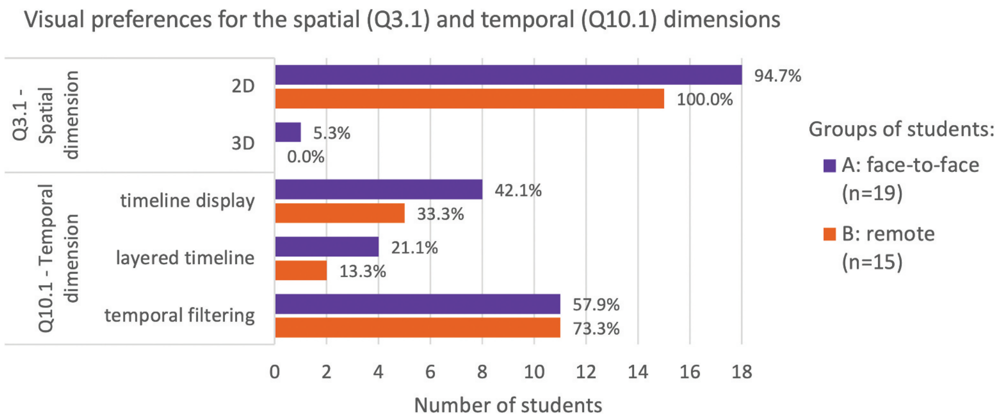

Figure 2. Summary of visual preferences for the spatial (Q3.1) and temporal (Q10.1) dimensions. Percentages are calculated for each group and question individually.

exemplified in Figure 1(c), where a visualization of the 16th, 19th, and 20th centuries is shown. Additionally, it should be noted that the layered timeline involves a 3D representation, which is not the preferred option when choosing between 2D and 3D.

Perceived usability of the tool

To assess the usability of the tool, we performed an SUS test (Brooke, 1996), which consisted of a questionnaire with ten questions (Appendix C). Participants rated each question (from SUS1 to SUS10) with a score in the range of 1–5, which means: 1: strongly disagree; 5: strongly agree. Odd questions (SUS1, SUS 3, . . . , SUS 9) are positive, so the best possible answer would be 5 points; on the other hand, even questions (SUS 2, SUS 4, . . . , SUS 10) are formulated in a negative way, so the best possible answer is 1 point. For this reason, the given scores by users are given in different graphs: in Figure 3, the positive questions; in Figure 4, the negative questions. In both cases, the best possible answers are depicted in magenta colour ("5: strongly agree" in Figure 3; "1: strongly disagree" in Figure 4), while the worst possible answers are in green ("1: strongly disagree" in Figure 3; "5: strongly agree" in Figure 4). As it can be seen, for the positive questions, the one receiving the maximum score (5 points) more times is SUS3 (A: $n = 6$; B: $n = 3$), that refers to the easiness of using the tool. SUS5 has similar results (5 points, A: $n = 6$; B: $n = 2$), which indicates that the integration of the system works well. The poorest results are for SUS1, as none of the students rated this question with the maximum score. However, this is understandable, as the question specifically asks users to assess whether they would use this system frequently, which is not the case, as the tool is focused on the field of silk heritage, which is a very specific field not directly related to the students' profile.

For the negative questions, the one receiving the best possible score (1 point) more times was SUS4 (A: $n = 10$; B: $n = 3$), indicating that the students would not need the support of a technical person to be able to use the tool and which is indeed aligned with the positive question SUS3 (easiness to use the tool) that also received the best possible score most times. On the other hand, question SUS6 received the poorest result in terms of how many students gave the best possible score (1 point, A: $n = 6$; B: $n = 0$). This

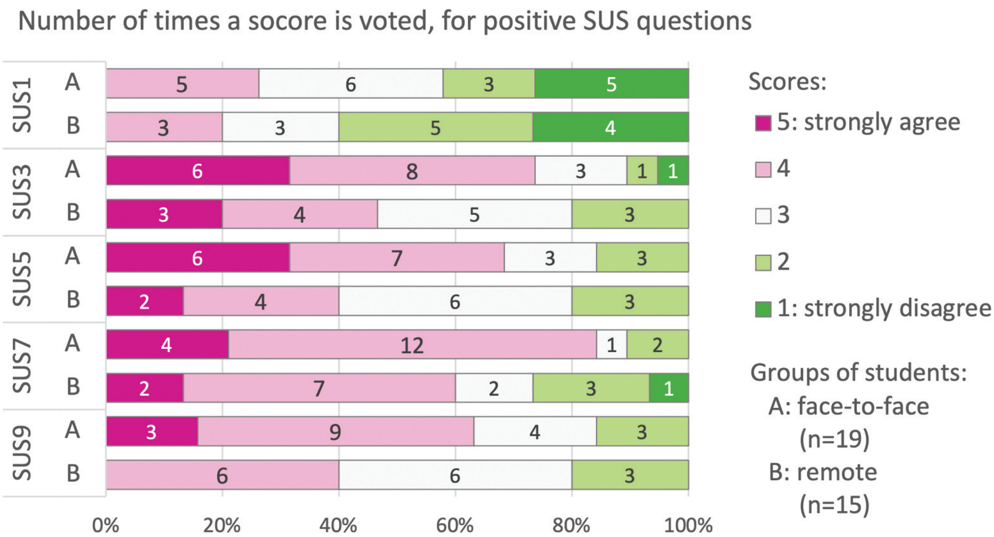

Figure 3. Number of times a score (from 1 to 5) is voted, given by the two groups (A: face-to-face; B: remote) for each positive SUS question, where: SUS1. I think that I would like to use this system frequently; SUS3. I thought the system was easy to use; SUS5. I found the various functions in this system were well integrated; SUS7. I would imagine that most people would learn to use this system very quickly; SUS9. I felt very confident using the system.

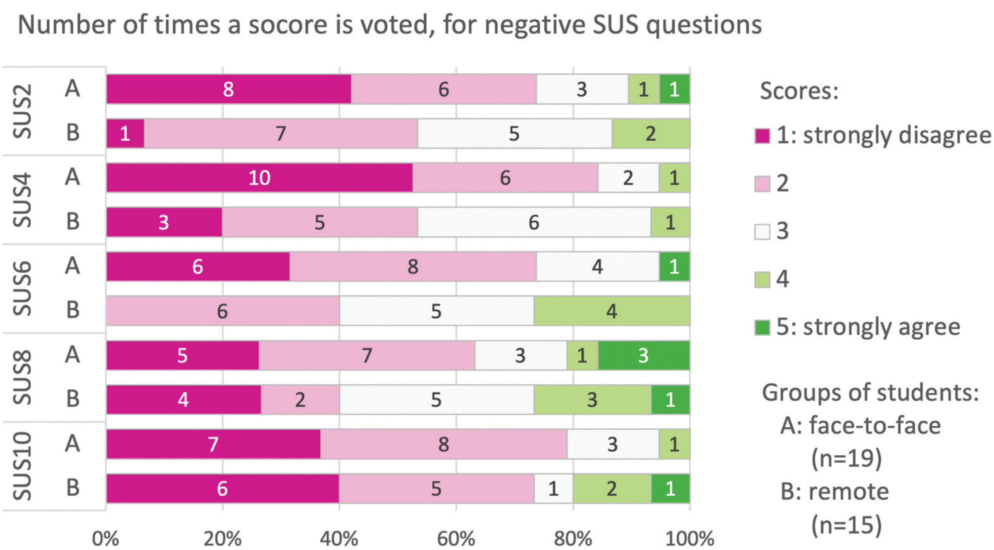

Figure 4. Number of times a score (from 1 to 5) is voted, given by the two groups (A: face-to-face; B: remote) for each negative SUS question, where: SUS2. I found the system unnecessarily complex; SUS4. I think that I would need the support of a technical person to be able to use this system; SUS6. I thought there was too much inconsistency in this system; SUS8. I found the system very cumbersome to use; SUS10. I needed to learn a lot of things before I could get going with this system.

Table 8. Results of the SUS test.

SUS Domain	Group A (sd)	Group B (sd)	All (sd)
SUS1	2.58 (1.17)	2.33 (1.11)	2.47 (1.13)
SUS2	2.00 (1.15)	2.53 (0.83)	2.24 (1.05)
SUS3	3.89 (1.10)	3.47 (1.06)	3.71 (1.09)
SUS4	1.68 (0.89)	2.33 (0.90)	1.97 (0.94)
SUS5	3.84 (1.07)	3.33 (0.98)	3.62 (1.04)
SUS6	2.05 (1.03)	2.87 (0.83)	2.41 (1.02)
SUS7	3.95 (0.85)	3.40 (1.18)	3.71 (1.03)
SUS8	2.47 (1.39)	2.67 (1.29)	2.56 (1.33)
SUS9	3.63 (0.96)	3.20 (0.77)	3.44 (0.89)
SUS10	1.89 (0.88)	2.13 (1.30)	2.00 (1.07)
SUS-Score	69.47 (15.96)	58.00 (16.26)	64.41 (16.87)

Table 9. Unpaired t-test and Mann-Whitney U tests for the SUS-Score.

Test	Group	Mean ± SD	F (sig.)	T (sig.)
Unpaired t-test	A	69.474 ± 15.9552	.257 (.616)	2.060 (.048)**
	B	58.000 ± 16.2624		
Test	Group	Median ± IQR	U (sig.)	
Mann-Whitney U	A	72.500 ± 17.500	77.500 (.023)**	
	B	57.500 ± 25.000		

Degrees of freedom = 32 (t-test).

question refers to inconsistencies in the tool. None of the students in group B rated it with 1 point, which is not the case for students in group A.

Finally, comparing Figures 3 and 4, overall, negative questions are better rated – in terms of the highest scores – than positive questions.

However, we can also analyse the results using the obtained SUS score instead of individual questions. The score ranges from 0 to 100, indicating that 100 is the best imaginable result. This score is depicted in Table 8, that also shows the average and standard deviation (sd) for each individual question and the SUS score.

For group A, the SUS score reaches 69.47 points, which can be considered good on the scale of scores provided by the questionnaire, and almost acceptable in relation to the acceptability ranges (Bangor et al., 2009; Brooke, 2013). However, the score given by group B was 58.00 points, which was significantly lower, as seen in Table 9, where an unpaired t-test and a Mann-Whitney U test to compare the SUS score of both groups are depicted. Both tests clearly showed that the SUS score of group A was significantly higher ($p < 0.05$) than the SUS score of group B.

Figure 5 shows the mapping of individual SUS scores for both groups in relation to the acceptability ranges, where discrepancies between the two groups can be observed. While most of the students in group A rated the system as acceptable ($n = 12$, 63.16%), most of the students in group B rated the system as marginal-low ($n = 5$, 33.33%).

Perceived efforts

The mean results of the NASA-TLX test (Appendix B) and their respective standard deviations (sd) are depicted in Table 10. The lower the value on the 1–21 scale the better is the result. The bold value indicates a lower result for a certain NASA-TLX question

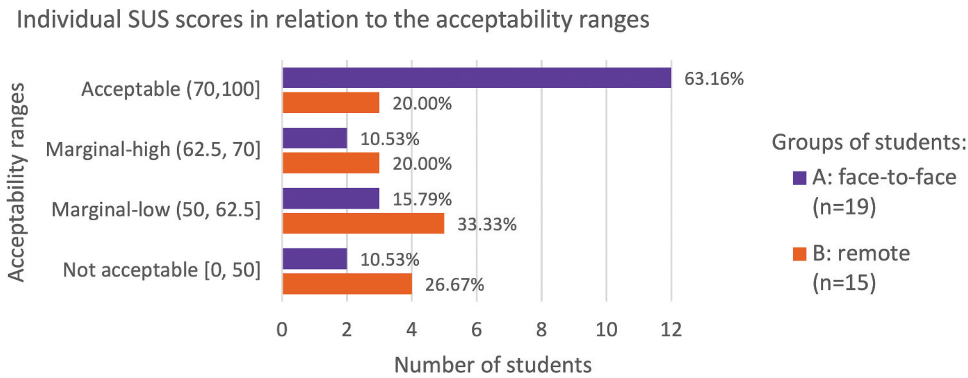

Figure 5. Individual SUS scores in relation to the acceptability ranges for the two groups. Percentages are calculated in relation to each group (a and b).

Table 10. Summary of the results of the NASA-TLX test.

NASA TLX Domain	Group A (sd)	Group B (sd)	All (sd)
1-Mental	**6.00** (7.79)	8.07 (4.74)	6.91 (4.04)
2-Physical	**4.11** (3.16)	5.40 (5.08)	4.68 (4.21)
3-Temporal	**8.16** (3.59)	10.73 (4.42)	9.29 (4.35)
4-Achievement	**4.21** (5.70)	6.27 (3.73)	5.12 (3.40)
5-Effort	**6.89** (5.47)	7.87 (4.19)	7.32 (4.15)
6-Frustration	**5.00** (5.65)	7.73 (5.81)	6.21 (4.84)
Avg	5.73 (5.23)	7.68 (4.66)	6.59 (4.16)

Table 11. Unpaired t-test for the average NASA-TLX score.

Test	Group	Mean ± SD	F (sig.)	T (sig.)
Unpaired t-test	A	5.73 ± 2.01	7.465 (.01)***	−2.078 (.046)**
	B	7.68 ± 3.42		

Degrees of freedom = 32 (t-test).

when comparing the two groups. Although all mean values for group A were lower, none of these six individual NASA-TLX questions provided results that were significantly better for either group under an unpaired t-test or a Mann-Whitney U test. For the sake of brevity, we do not provide the detailed results of these tests. However, the average NASA-TLX Scale showed a significantly better value for group A under a parametric t-test (p-value of 0.046), as shown in Table 11. Therefore, we can conclude that the students in group A perceived the tasks as less demanding.

Qualitative study

As mentioned above, an inductive coding methodology was used to analyze the students' responses. Regarding question Q3.1.1, the results are summarized in Figure 6 (for the two groups together, $N = 34$) and Figure 7 (for groups A and B). As can be seen in Figure 6, 35% ($n = 12$, from $N = 34$) preferred 2D for better visualization, that is, those who considered that "the map looks better" or "it is better visualized". In this category we also considered those who referred to markers as "good". Also, a 26% ($n = 9$), preferred

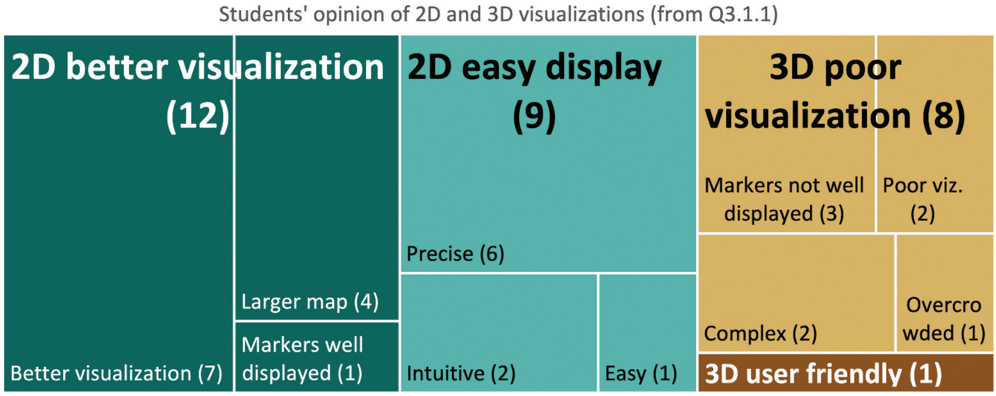

Figure 6. Tree map showing the students' likeness and unlikeness of 2D and 3D, after analyzing responses to Q3.1.1. The numbers of students are shown in parentheses.

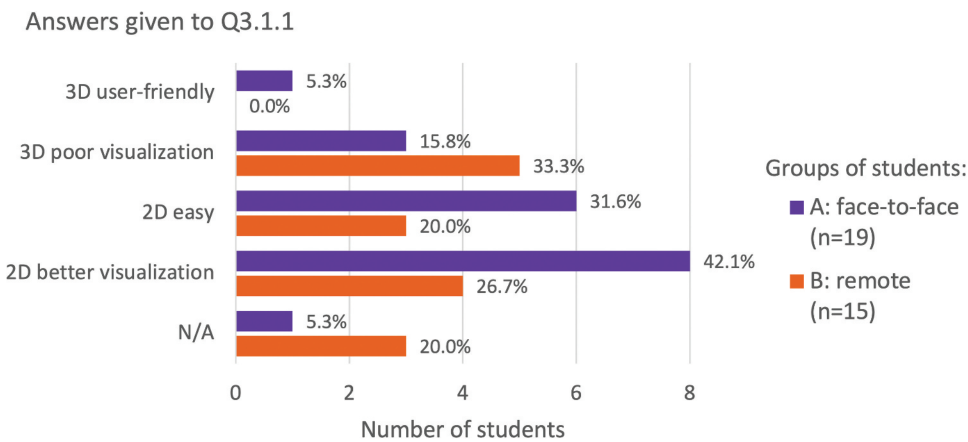

Figure 7. Graphic comparing the answers to Q3.1.1 for the two groups. Percentages are calculated in relation to each group (a and b).

2D for its ease of use, i.e. when students mentioned that it was "intuitive", "easy to use", or "better understood". On the other hand, we found that only one person liked 3D and the rest of the comments are negative, 24% ($n = 8$) in relation to the poor visualization of 3D ("bad visualization"; "no good markers"; "visualization problems"= "poor visualization"; "no good markers"; "visualization problems"). Those who preferred 2D for better visualization, the majority ($n = 7$) directly allude to this aspect, while the rest prefer it for allowing the map to be seen more extensively ($n = 4$) and the fact that the markers can be seen better ($n = 1$). In contrast, among those who think that the 3D visualization is bad, some think that it prevents them from seeing the markers correctly ($n = 3$), others think that the visualization is poor and complex ($n = 2$), and one thinks that the visualization is overcrowded ($n = 1$).

When comparing the two groups, as shown in Figure 7, we observed that, overall, both groups liked more 2D visualization, including a wider view of the maps and better markers, while they rated the 3D visualization negatively because it was more overcrowded, making it more complex to view. However, there are some differences between them. It seems that group A tends to give responses in a positive way; for example, they prefer 2D because either they find the visualization better or it is easy to understand, while group B tends to give responses in a negative way, for example, they prefer 2D because 3D gives a poorer visualization. Additionally, it is worth mentioning the greater number of non-available answers for group B ($n = 3$, 20.0%) in comparison to group A ($n = 1$, 5.3%).

Regarding the temporal scale, in Q10.1.1, participants were asked to explain why they liked their selection. As in the previously analyzed question, in some cases, they answered in relation to what they did not like about other possibilities. Therefore, we decided to analyze the two questions independently. The results are summarized in Figure 8 (for the two groups together, $N = 34$) and Figure 9 (for groups A and B). As shown in Figure 8, students preferred temporal filters ($n = 14$) because they were intuitive, allowed better

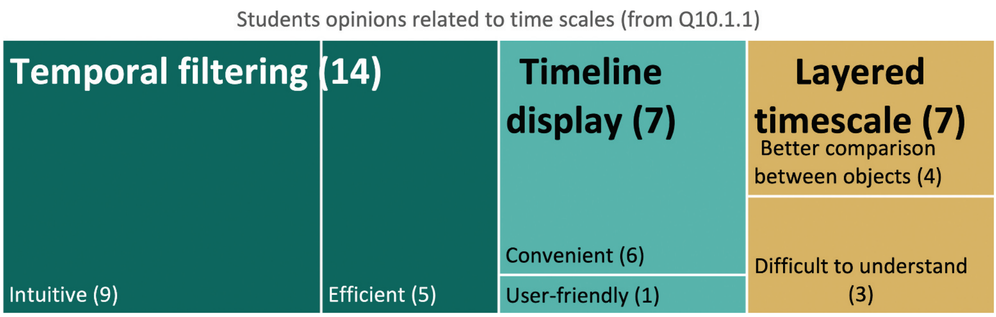

Figure 8. Tree map showing the students' opinions related to time scales, after analyzing responses to Q10.1.1. In parenthesis, the number of students is shown.

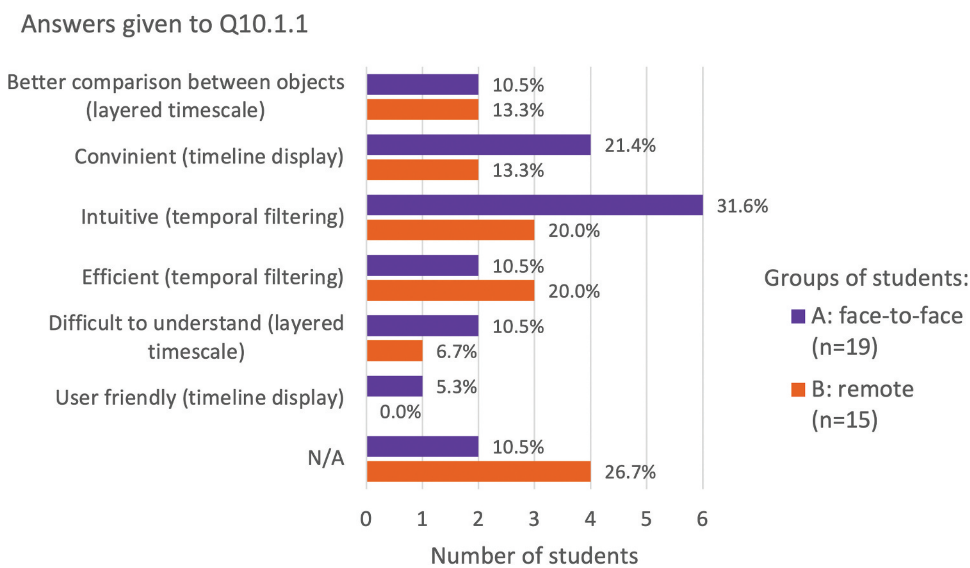

Figure 9. Graphic comparing the answers given to Q10.1.1 for the two groups. Percentages are calculated in relation to each group (a and b).

visualization, and were more efficient. On the other hand, they find line-based relationships convenient and easy to use. It is striking that practically the same number of students found that stacked century maps allow a better comparison between objects ($n = 4$) and others that are difficult to understand ($n = 3$). Both the temporal filtering and timeline displays had positive opinions. For the first, students found it intuitive ($n = 9$) and efficient ($n = 5$), while the timeline display was convenient ($n = 6$) and user-friendly ($n = 1$). On the contrary, the layered timescale had both positive and negative opinions; while students found that this type of display allows for better comparison of objects ($n = 4$), others found it difficult to understand what they were looking at ($n = 3$).

Finally, comparing both groups as shown in Figure 9, we observe that the two of them responded in a similar way to the reasons for their preferences in the selection of the time filter; however, similar to the previous analyzed question, here we observe a greater number of non-available answers for group B ($n = 4$, 26.7%) in comparison to group A ($n = 2$, 10.5%).

Discussion

Improving and maximizing student engagement and motivation in both in-person and remote online learning modalities requires tailored approaches that recognize the strengths and limitations of each environment. In this paper, we have conducted a workshop in a hybrid environment were both modalities were considered, adding complexity to the planned activities, as they had to be appropriated for the two modalities.

The accomplishment of the first aim of the research, that is, to transfer the research outcomes to the classroom, was straightforward, simply fulfilled from the organization of the workshop that included both the initial explanation of the project and the realization of specific tasks with the STMaps tool, which is one of the project's outcomes. As most of the students were able to complete the tasks and respond well to questions that required a specific answer (average success rate: 83.46% for group A; 80.00% for group B), we can conclude that the transfer of the research outcomes to the classroom was properly accomplished. Additionally, from other analyzed parameters, the visual aspects of the spatial and temporal dimensions were specifically related to the subject of data visualization, providing the students with extended options in comparison to what they had seen in the subject so far, for instance, the possibility of switching between 2D and 3D visualization, the 3D stacked maps for the temporal inspection of objects, the visualization of objects relationships, etc. Thus, students benefited from experimenting with ways of visualizing data that were complementary to other contents of the subject.

Regarding the second aim to transfer back students' feedback to the research activity, the participation of the technical staff of the project during the workshop was essential, as they were able to observe how students managed the tool and the difficulties they faced in fulfilling certain tasks. From the analyzed parameters, the feedback of the visual aspects was somewhat relevant, not so much for the STMaps tool itself, as it was in the last stages of development, and it was decided not to include significant changes, but for other research activities carried out by the same team. Specifically, another spatiotemporal map developed by the team (Portalés et al., 2022), from its design, included aspects that were found relevant after the analysis of STMaps, for example, 3D visualization and layered maps were not considered, the filtering menu was simplified in terms of graphical interface, and different strategies were implemented according to the number of objects in a single cluster. On the other hand, the acceptable results of both perceived usability and perceived efforts were relevant to validate STMaps during the project lifetime, which is specifically one step considered in the research methodology, thus having a direct impact on the project results.

For the third aim, that is to assess the quality of the learning process in a hybrid classroom, interesting conclusions can be drawn from almost all the studied parameters. First, when analyzing correctness, there were no significant differences between the two groups. Thus, we cannot state that students in the classroom performed better than those at home. However, derived from both the motivation parameter and the qualitative analysis, the results suggest that those at home were less motivated, responded to fewer questions (when explanations were required), and, in the specific case of justifying if they preferred 2D or 3D visualization, tended to give answers in a negative way. Additionally, when analysing the individual SUS scores per each group, it is relevant to see significant differences, allowing us to think that one reason could be that group A was more receptive and motivated towards the workshop than group B, so group B penalized the tool. In addition, it is relevant to see the differences in the perceived efforts, with group B experiencing slightly more difficulties in all the six parameters measured with the TLX scale; however, as the correctness parameter indicates that both groups performed similarly, this might also be an effect of penalizing the tool with lower scores because of being less motivated. A possible limitation of the workshop in relation to group B was that students were not required to switch their camera on, what might have influenced the motivation parameter in a negative way.

In a different line of thought, it is also essential to recognize the relationship between data visualization and geography must be also acknowledged. Geography, as a broad discipline, encompasses a wide range of sub-disciplines, such as human geography, physical geography, people-environment interactions, and cartography/GIScience. The relevance of data science and data visualization within these areas may vary depending on the specific focus. For instance, in fields like cartography and GIScience, the connection to data science and visualization is more explicit, as these areas rely heavily on spatial data analysis and the creation of visual representations of geographic information. In contrast, sub-disciplines such as human geography or people-environment studies may have a less direct but still significant connection to data science. For example, researchers in these fields may utilize data visualization techniques to illustrate social patterns, environmental changes, or migration flows, helping to convey complex spatial relationships. Therefore, we believe that educators across all areas of geography should recognize the importance of data visualization in enhancing students' ability to analyze and communicate geographic-related information effectively.

Conclusions

In this paper we have presented the results of a workshop carried out with students of the second year of a university degree on data science, that took place in the context of the pandemic restrictions, and therefore some of the students were in the classroom (group A, $n = 19$, 55.88%) and the others were at home (group B, $n = 15$, 44.12%). They learned how to use a digital tool named STMaps, that has been developed in the scope of the

SILKNOW project. This let us transfer research outcomes to the classroom as well as transfer back students' feedback to the research activity; a win-win transfer of knowledge. But more importantly, because of the special context of the workshop, i.e. a hybrid learning environment, we could assess if the learning process of students that were online was of the same quality as of those that were physically at the classroom.

During the workshop, students were asked to carry out some tasks and complete a questionnaire. Then, we analyzed the responses in terms of the correctness, motivation, visual preferences, perceived usability, and perceived effort, and further performed a qualitative analysis. The results of our experiment show that there are some differences in the learning process in the two groups. While both performed well in terms of correctly fulfilling the tasks, we have insights that the students at home were less motivated and perceived the tasks more cumbersome to achieve that the ones present in the classroom.

We hold the view that, while at the academic level, there is little difference in using these technologies in face-to-face and e-learning settings, attending an in-person workshop can enhance student satisfaction and motivation. This is because there is usually a higher level of commitment, and direct interaction with instructors and peers is highly valued. This bidirectional knowledge-transfer activity was perceived as very useful, and similar activities are being planned for each academic year.

Following the completion of this research, and concurrently with the knowledge acquired from various similar endeavours, the authors of this study posit that the inclusion of students in the planning and design of research projects, as recommended by Felten & Clayton (2011), is highly advantageous in terms of strategic and proactive planning. This approach is facilitated by the utilisation of technological tools that are characterised by an innovative element and other relevant factors, such as proximity to a more professional environment. Furthermore, in accordance with the suggestions of Brew and Saunders (2020), this approach assists students in developing research skills, training in data analysis, and proficiency in communication tasks. Lastly, the implementation of interdisciplinary methodologies is of paramount importance, as it enhances the quality of the learning experience.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The research leading to these results is in the frame of the "SILKNOW. Silk heritage in the Knowledge Society: from punched cards to big data, deep learning and visual/tangible simulations" project, which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 769504. Cristina Portalés is supported by the Spanish government postdoctoral grant Ramón y Cajal, under grant No. RYC2018-025009-I.

ORCID

Cristina Portalés http://orcid.org/0000-0002-4520-2250

Compliance with ethical standards

All the participants in this research volunteered for the experiment and gave the consent regarding their participation in the study, which was completely anonymous.

Data availability statement

The datasets analysed during the current study are available in the project repository, at https://silknow.eu/silknow/STMAPS_Evaluation/.

References

Aguilera-Hermida, A. P. (2020). College students' use and acceptance of emergency online learning due to COVID-19. *International Journal of Educational Research Open*, *1*, 100011. https://doi.org/10.1016/j.ijedro.2020.100011

Alvarez-Herrero, J.-F., & Hernández Ortega, J. (2020). Evaluación y rendimiento académico del alumnado universitario en formación online versus formación presencial. In E. Colomo Magaña (Ed.), *La tecnología como eje del cambio metodológico* (pp. 1136–1139). UMA Editorial.

Bailey, L. M. (2020). *Comparing students' learning outcomes and satisfaction in online, hybrid and face-to-face education courses*. Ed.D. Temple University.

Bandura, A. (Ed). (1995). *Self-efficacy in changing societies* (1st ed.). Cambridge University Press.

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, *4*(3), 114–123.

Bond, M. (2020). Facilitating student engagement through the flipped learning approach in K-12: A systematic review. *Computers & Education*, *151*, 103819. https://doi.org/10.1016/j.compedu.2020.103819

Bond, M., Marín, V. I., Dolch, C., Bedenlier, S., & Zawacki-Richter, O. (2018). Digital transformation in German higher education: Student and teacher perceptions and usage of digital media. *International Journal of Educational Technology in Higher Education*, *15*(1), 48. https://doi.org/10.1186/s41239-018-0130-1

Bower, M. (2019). Technology-mediated learning theory. *British Journal of Educational Technology*, *50*(3), 1035–1048. https://doi.org/10.1111/bjet.12771

Bozkurt, A., & Sharma, R. C. (2020). Emergency remote teaching in a time of global crisis due to CoronaVirus pandemic. *Asian Journal of Distance Education*. https://doi.org/10.5281/ZENODO.3778083

Brew, A., & Saunders, C. (2020). Making sense of research-based learning in Teacher education. *Teach. Teach. Edu*, *87*, 102935. https://doi.org/10.1016/j.tate.2019.102935

Brooke, J. (1996). SUS-A quick and dirty usability scale. Usability evaluation in industry 189.194.

Brooke, J. (2013). SUS: A retrospective. *Journal of Usability Studies*, *8*, 29–40.

Capone, R., De Caterina, P., & Mazza, G. (2017). Blended learning and flipped learning: challenges and opportunities for the 21st-Century students to create a digital environment. *Digital Education for the 21st Century*. 10478–10482.

Conelly, J., & Miller, P. (2018). Improving learning outcomes for higher education through smart technology. *International Journal of Conceptual Structures and Smart Applications (IJCSSA)*, *6*(1), 1–17. https://doi.org/10.4018/IJCSSA.2018010101

Daniela, L., Visvizi, A., Gutiérrez-Braojos, C., & Lytras, M. (2018). Sustainable higher education and technology-enhanced learning (TEL). *Sustainability*, *10*(11), 3883. https://doi.org/10.3390/su10113883

Davies, R. S. (2011). Understanding technology literacy: A framework for evaluating educational technology integration. *Techtrends Tech Trends*, *55*(5), 45–52. https://doi.org/10.1007/s11528-011-0527-3

Dunn, T. J., & Kennedy, M. (2019). Technology enhanced learning in higher education; motivations, engagement and academic achievement. *Computers & Education*, *137*, 104–113. https://doi.org/10.1016/j.compedu.2019.04.004

EDUCAUSE. (2018). *Report from the 2018 EDUCAUSE task force on digital transformation*.

Elfaki, N. K., Abdulraheem, I., & Abdulrahim, R. (2019). Impact of e-learning vs traditional learning on student's performance and attitude. *International Journal of Medical Research & Health Sciences*, *8*(10), 76–82.

Faura-Martínez, U., Lafuente-Lechuga, M., & Cifuentes-Faura, J. (2021). Sustainability of the Spanish university system during the pandemic caused by COVID-19. *Educational Review*, *74*(3), 645–663. https://doi.org/10.1080/00131911.2021.1978399

Felten, P., & Clayton, P. (2011). Service-learning. *New Directions for Teaching and Learning*. https://doi.org/10.1002/tl.470

Halasa, S., Abusalim, N., Rayyan, M., Constantino, R. E., Nassar, O., Amre, H., Sharab, M., & Qadri, I. (2020). Comparing student achievement in traditional learning with a combination of blended and flipped learning. *Nursing Open*, *7*(4), 1129–1138. https://doi.org/10.1002/nop2.492

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). North Holland Press.

He, Z., & Schonlau, M. (2020). Automatic coding of open-ended questions into multiple classes: Whether and how to use double coded data. *Survey Research Methods*, 267–287. https://doi.org/10.18148/SRM/2020.V14I3.7639

Jenkins, A. (2000). The relationship between teaching and research: Where does geography stand and deliver? *Journal of Geography in Higher Education*, *24*(3), 325–351. https://doi.org/10.1080/713677414

Kemp, A., Palmer, E., & Strelan, P. (2019). A taxonomy of factors affecting attitudes towards educational technologies for use with technology acceptance models. *British Journal of Educational Technology*, *50*(5), 2394–2413. https://doi.org/10.1111/bjet.12833

Kurniawan, Y., Yulanthari, C., Ampow, M., Prahastuti, M., Anwar, N., & Cabezas, D. (2022). Evaluation of hybrid learning in the university: A case study approach. *HighTech and Innovation Journal*, *3*(4), 394–410. https://doi.org/10.28991/HIJ-2022-03-04-03

Kyngäs, H. (2020). Inductive content analysis. In H. Kyngäs, K. Mikkonen, & M. Kääriäinen (Eds.), *The application of content analysis in nursing science research* (pp. 13–21). Springer International Publishing.

Moore, S., Haviland, D., Moore, W., & Tran, M. (2016). Preparing teachers to use GIS: The impact of a hybrid professional development program on teachers' use of GIS. *Journal of Science Education and Technology*, *25*(6), 930–946. https://doi.org/10.1007/s10956-016-9641-5

Pal, D., & Vanijja, V. (2020). Perceived usability evaluation of Microsoft teams as an online learning platform during COVID-19 using system usability scale and technology acceptance model in India. *Children & Youth Services Review*, *119*, 105535. https://doi.org/10.1016/j.childyouth.2020.105535

Portalés, C., Casanova-Salas, P., Sevilla, J., Sebastián, J., León, A., & Samper, J. J. (2022). Increasing access to cultural heritage objects from multiple museums through semantically-aware maps. *ISPRS International Journal of Geo-Information*, *11*(4), Article 4. 266. https://doi.org/10.3390/ijgi11040266

Portalés, C., Sebastián, J., Alba, E., Sevilla, J., Gaitán, M., Ruiz, P., & Fernández, M. (2018). Interactive tools for the preservation, dissemination and study of silk heritage—an introduction to the SILKNOW project. *Multimodal Technologies and Interaction*, *2*(2), 28. https://doi.org/10.3390/mti2020028

Safwa, B., Sayed, A., & AlMetwaly, W. (2022). Enhancing hybrid learning using open source GIS-Based maps archiving system. *Egyptian Journal of Remote Sensing and Space Science*, *25*(3), 779–793. https://doi.org/10.1016/j.ejrs.2022.07.003

Sevilla, J., Casanova-Salas, P., Casas-Yrurzum, S., & Portalés, C. (2021). Multi-purpose ontology-based visualization of spatio-temporal data: A case study on silk heritage. *Applied Sciences*, *11*(4), Article 4. 1636. https://doi.org/10.3390/app11041636

Soffer, T., & Nachmias, R. (2018). Effectiveness of learning in online academic courses compared with face-to-face courses in higher education. *Journal of Computer Assisted Learning*, *34*(5), 534–543. https://doi.org/10.1111/jcal.12258

Tang, T., Abuhmaid, A. M., Olaimat, M., Oudat, D. M., Aldhaeebi, M., & Bamanger, E. (2020). Efficiency of flipped classroom with online-based teaching under COVID-19. *Interactive Learning Environments*, *31*(2), 1077–1088. https://doi.org/10.1080/10494820.2020.1817761

Thai, N. T. T., De Wever, B., & Valcke, M. (2017). The impact of a flipped classroom design on learning performance in higher education: Looking for the best "blend" of lectures and guiding questions with feedback. *Computers & Education*, *107*, 113–126. https://doi.org/10.1016/j.compedu.2017.01.003

Vallée, A., Blacher, J., Cariou, A., & Sorbets, E. (2020). Blended learning compared to traditional learning in medical education: Systematic review and meta-analysis. *Journal of Medical Internet Research*, *22*(8), e16504. https://doi.org/10.2196/16504

Verhagen, E., Voogt, N., Bruinsma, A., & Finch, C. F. (2014). A knowledge transfer scheme to bridge the gap between science and practice: An integration of existing research frameworks into a tool for practice. *British Journal of Sports Medicine*, *48*(8), 698–701. https://doi.org/10.1136/bjsports-2013-092241

Xing, B., & Marwala, T. (2017). Implications of the fourth industrial age for higher education. The Thinker 73: *Tạp chí Nghiên cứu dân tộc*, (23). https://doi.org/10.25073/0866-773X/87

Zhang, L., Guo, Y., & Zhang, Y. (2022). Discussion on GIS experimental hybrid teaching method for surveying and mapping engineering. *Bulletin of Surveying & Mapping*, *0*(2), 164–167,173. https://doi.org/10.13474/j.cnki.11-2246.2022.0064