AGILE: GIScience Series, 4, 13, 2023. https://doi.org/10.5194/agile-giss-4-13-2023
Proceedings of the 26th AGILE Conference on Geographic Information Science, 2023.
Editors: P. van Oosterom, H. Ploeger, A. Mansourian, S. Scheider, R. Lemmens, and B. van Loenen.
This contribution underwent peer review based on a full paper submission.
© Author(s) 2023. This work is distributed under the Creative Commons Attribution 4.0 License.

Does spatial thinking ability relate to performance when using web-mapping services? A survey with digital natives

Eleni Tomai[1], Margarita Kokla[1], Christos Charcharos[1], and Marinos Kavouras[1]

[1]School of Rural, Surveying and Geoinformatics Engineering, National Technical University of Athens, Athens, Greece

Correspondence: Eleni Tomai (etomai@mail.ntua.gr)

Abstract. We present the design and implementation of an empirical synchronous remote study for exploring the relation between spatial ability and performance on web-mapping services involving undergraduate University students; digital natives. The study exploits Spatial Thinking Ability Test to assess participants' spatial ability and to reveal if and to what extent it is related to their ability to perform tasks on popular web-mapping services. Participants' performance was assessed on the basis (a) of successfully executing tasks and (b) of how much time participants needed to properly perform tasks. A usability scale was used to measure participants' subjective perceptions of web-mapping services usability. Moreover, participants were self-assessed in digital skills using the Digital Natives Assessment Scale. Results reveal differences among services in task accuracy indicating that not only the web-mapping service but also the nature of tasks guide participants' performance. Correlations between spatial ability, digital skills, system usability, familiarity, and performance using web mapping services tend to be low and not significant leading to the assumption that success when interacting with a web mapping service is underlined by other factors as well.

Keywords. Spatial ability, web mapping service, digital skills, system usability, familiarity, digital natives

1 Introduction

The paper describes the design and implementation of an empirical study, conducted on-line, to explore whether young people in their twenties with "engrained" digital skills; digital natives (Prensky, 2001) and with basic geospatial educational background interact fruitfully with web mapping services, in other words, if there is a transfer of digital skills and of spatial ability to map using skills.

Three well known web-based mapping services (Google Maps - GM, Bing Maps - BM, and HERE WeGo - HWG) were used to measure participants' performance along two axes (a) accuracy; measuring whether participants correctly performed the tasks, (b) efficiency; measuring the time participants needed to correctly perform them. Participants' subjective perceptions of the services' usability (user satisfaction) were measured, through the System Usability Scale (SUS; Brooke, 1996). Participants' spatial ability was assessed using Spatial Thinking Ability Scale (Lee and Bednarz, 2012). Additionally, participants were asked to self-assess their digital skills by responding to the Digital Natives Assessment Scale (DNAS, Teo, 2013). They also stated their familiarity with each one of the three web mapping services used in the survey.

Similar line of research regarding the effect of spatial ability on map learning has been explored by Çöltekin et al. (2018), while the effect of map resolution and of spatial abilities on map learning has also been studied (Sanchez and Branaghan, 2009).

Web mapping services are nowadays used by a vast number of users in their everyday life. The explosion of their use is attributed to the interactivity, support for collaboration, and ease of use of 2D and 3D mapping (Veenendaal et al., 2017).

Usability testing is a widely used method for the evaluation of systems, products or services. Usability is defined as the "extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (ISO, 2018) Effectiveness is defined as the accuracy and completeness with which users achieve specified goals, while efficiency is defined in terms of resources (e.g., time, human effort, costs and material) used in relation to the results achieved.

Satisfaction indicates the extent to which the user experience and responses that result from the use of a system, product or service meet the user's needs and expectations (ibid).

Regarding the testing environment, usability tests have been typically performed in laboratory conditions (in-lab) involving a small number of representative users who performed tasks using a service or prototype. However, issues related to the limited number of participants tested, as well as the high cost required for testing have led to the development of remote testing techniques (Tullis et al., 2002; Alghamdi et al., 2013). This study followed a synchronous remote testing that involved communication between test administrators and participants in real time via technical means.

The paper is organized as follows. Section 2 presents related work on STAT, DNAS, and usability of web mapping services. Section 3 details the study design and process. Section 4 presents the results of the statistical analysis, while the last section discusses the findings.

2 Related Work

2.1 Spatial Ability Assessment using STAT

Vast volumes of research results on spatial abilities/ skills assessment exploit batteries of psychometric tests, many developed in the 70's, which adopt narrow and fragmented views of spatial abilities, whereas nowadays spatial ability is considered an encompassing cognitive ability than cross-cuts different facets of human cognition.

Spatial Thinking Ability Test (STAT) (Lee and Bednarz, 2012) is an instrument that can measure spatial ability, while integrating geography content knowledge and spatial skills. The test has been initially validated by a sample of 532 students of different educational levels; junior high, high school, and university (ibid.)

STAT has been used to assess spatial thinking of university and high school students by researchers in different countries worldwide (Bednarz & Lee, 2019). It is an easy to use, adaptable to different cultural environments and flexible to different settings making it a suitable measure of assessing spatial thinking for the current study, especially since participants have already been exposed to concepts of space and spatial analysis, included in STAT, via their field of studies (section 3.2).

2.2 Digital Natives Assessment Scale

DNAS (Teo, 2013) was initially developed to identify whether and to what extent individuals assess themselves as being digital natives. DNAS consists of four factors which seem to be sufficiently distinct and can encompass the essence of being a digital native; (1) grow up with technology, (2) comfortable with multitasking, (3) reliant on graphics for communication, and (4) thrive on instant gratifications and rewards.

It was validated by 1018 students and has been used in the education sector in exploring students' interactions with technology and digital exposure and use.

2.3 Web-mapping Services Usability Assessment

We detail some studies of web mapping services usability testing in chronological order.

Skarlatidou and Hakley (2006) organized two workshops to perform usability testing of web mapping sites (participants performed 6-7 tasks). The first evaluated Multimap, Google Maps and Map Quest, whereas the second, MSN maps, Yahoo! (European) maps, ViaMichelin and StreetMap. The think aloud method, a pre-test and a post-test questionnaire were used to gather qualitative data. Quantitative data included completion time for each task, total number of clicks, and success rate. Qualitative results showed important usability issues related to the website design, the maps size and design, the symbols used, and the functionality offered.

Wachowicz et al. (2008) proposed a usability framework for web mapping services of five abstraction levels (hypothesis, typology, variables, elements, and measures), each providing the theoretical basis for the design and implementation of usability tests to measure user satisfaction of web mapping services. An empirical study involving two commercial web mapping route planning services and three tasks, implemented the proposed usability framework testing whether higher degree of usability of web mapping services can be associated with greater user satisfaction. User satisfaction was measured on the basis of: speed of performance, rate of interaction (number of interactions used to perform a task), and rate of error (number of errors when performing a task). The authors highlighted the difficulty in studying user satisfaction without taking into account other factors such as familiarity and operability.

Nivala et al. (2008) performed in-lab usability evaluation of Google Maps, MSN Maps & Directions, MapQuest, and Multimap involving 8 non-expert and 16 expert users (cartographers and usability engineers). Non-experts performed seven tasks and described the reasoning of their actions using the 'think aloud' method. Experts were given the list of tasks and were asked to record all problems encountered during the tasks' execution. The identified usability problems were grouped based on their severity. Based on the evaluation, the authors proposed design guidelines for the identified usability problems.

Khan and Adnan (2010) carried out in-lab comparative study of two well-known web mapping services; Google Maps and MapQuest. The evaluation involved six users (four experienced and two novices) who performed four tasks. The evaluation was based on completion time and observations on the use of the two web mapping services. Users also responded to a questionnaire for evaluating parameters such as effectiveness, usefulness, user reaction, consistency, visual clarity, and functionality and to open-ended questions to further explain their opinions of the services used.

Wang (2014) explored usability problems of Google Maps, Bing maps, MapQuest, and Yahoo Maps. The in-lab study involved 42 users who performed five tasks and the whole process was recorded. The usability evaluation was measured based on: the task completion time, the whole time of the experiment for each participant, the total number of clicks, and the task success rate. A post-test questionnaire was used to identify errors and mapping services usability problems related to the user interface, the functionality, the search operations and the visualization aspects.

Kavouras et al. (2021) performed comparative usability analysis of four web mapping services; Google Maps, Bing Maps, Here WeGo, and MapQuest. 167 university students performed tasks on these services and results indicate no significant differences between male and female participants in terms of efficiency or effectiveness. Substantial part of participants' performance and interaction with web mapping services is mostly task- than gender-driven; a “difficult” task seems to be the overarching factor of participants' performance. Task “difficulty”, however may depend on the service’s usability issues.

Web mapping services have been typically evaluated using laboratory testing approaches involving limited number of users. Due to the widespread use of these services, especially by young people, a study involving a larger number of users interacting with these services in familiar and less artificial environments (e.g., home, work) where typically the tasks are performed in everyday life, is considered important to give further insights into significant usability aspects.

Thus, the present study took a different approach for evaluating the usability of three well-known web mapping services. A synchronous remote testing was conducted involving 74 young users who performed five tasks in their familiar environments. The test administrators acted as facilitators interacting with the participants in real time without observing their actions.

3 Study Design and Process

Participants filled in a demographics questionnaire followed by a session during which they answered to the STAT Questionnaire and DNAS. Then, they were split into two equally-numbered groups and they performed five tasks on two different web mapping services each group (three in total since Google Maps was used by both groups) followed by a closing session where they assessed each service’s usability. Participants were informed of all the steps of the study beforehand. Test administrators were available during the study in case participants had any questions on the process, on the questionnaires, or any problems in accessing the services. Before the survey, participants’ questions were answered to minimize any misunderstandings or misconceptions that could pose limitations on survey results. The entire survey questionnaire was created in Alchemer (2022).

3.1 Pre-test Questionnaire

The pre-test questionnaire was used to determine sample demographics, such as gender, age group, the reasons why participants use computers and smart devices (multiple choice possibility from a list of answers e.g., social networking, entertainment, education, work), web-mapping service familiarity (5-point Likert scale) and finally familiarity with the Boston area (5-point Likert scale), where the web mapping service use scenario takes place (section 3.5).

Finally, participants were asked on the kind of device they were intending to use to perform the survey. Only two answered they were going to use tablet or mobile phone. They were both prompted to change to laptop or desktop before moving on to the STAT and the Web Mapping Services questionnaires.

3.2 Sample Description

The sample consists of 74 undergraduate students (40 male) of the School of Rural, Surveying and Geoinformatics Engineering of the National Technical University of Athens, Greece. Student participation was consensual, voluntary, and anonymous without any personal data collection.

All participants belong to the 19-24 age group, but one, (25-30). Participants were enrolled, at the time of the survey, in the fourth semester, which includes courses such as Introduction to GIS and Analytical Cartography, of a five-year (10 semesters) study cycle. The year before, during the second semester, they had completed the Introduction to Cartography course. They have acquired some basic knowledge and training on maps, map symbols, legend and scale, cartographic design, GIS principles and theory, spatial queries, spatial analysis etc., and that they comprise a homogeneous group, so no differences in their level of geospatial knowledge exist that could affect the results.

3.3 STAT Questionnaire

STAT Questionnaire comes in two versions, comprised of 16 items each, their differences being in the order of multiple-choice options, and there are some figures inversions. For this study, we used version A. The questionnaire was translated into Greek. One master's and two PhD students, all in the field of GI Science, took the test to validate it. They were asked not only to solve it but also to evaluate the test language and terminology, the clarity and meaning of the text in Greek. The time it took them to complete the test was recorded to help the survey administrators determining the time limits to impose on the actual survey. Based on the three students comments the translation and terminology were fine-tuned for the final Greek online version.

3.4 Digital Natives Assessment Scale

DNAS consists of 21 items rated by a 7-point Linkert scale ranging from "strongly disagree" (1) to "strongly agree" (7). The scores are summed, thus ranging from 21 to 147, reflecting the degree of an individual's perceptions of being a digital native. DNAS items were translated into Greek. The same individuals who validated STAT, performed the similar endeavour of validating DNAS translated version.

3.5 Web Mapping Services Usability Questionnaire

The usability questionnaire consists of two sections. The first comprises the tasks that the participants had to perform using the web mapping services, while the second is the System Usability Scale (SUS) (Brooke, 1996).

3.5.1 Execution of Tasks

For this questionnaire section, two versions were created; each of them evaluating the performance of participants on a different pair of web mapping services. It was decided that all participants use Google Maps (GM), considered it is the most used mapping service worldwide and thus it can serve as a benchmark of their performance. The other two services were: Bing Maps (BM) and Here WeGo (HWG). The sample was equally and randomly divided into two groups who used the three services in pairs. The tasks that the participants had to perform were:

- identifying a location (point) using its geographical coordinates (Task 1),
- providing the postal address of a POI (Task 2),
- finding the shortest path (in terms of distance) between two points (Task 3),
- finding the shortest walking path (in terms of time) between two points (Task 4), and finally,
- locating the nearest point of interest to a specific point (Task 5).

For Task 5 participants were asked the follow-up multiple-choice question "How did you answer this question?", with three available answers: (1) by viewing/examining the area and by changing the zoom level (zooming in/ out), (2) by using a specific/ designated functionality offered by the service, (3) by other means; please elaborate. This provided an indication on how well they know different functionalities of the services in question.

The tasks have been chosen in line with the papers referred in the related work section (2.3). Specifically:

- In Nivala et al. (2008), a typical scenario of web mapping services use was presented to participants: "A tourist plans to visit London and uses a web mapping service to plan the trip in advance." Users had to perform seven tasks: (a) locating a point of interest (twice), (b) locating a point given a criterion, (c) finding a route, (d) selecting the most suitable location, (e) calculating a distance, (f) returning to a selected location.
- In Wachowicz et al. (2008), participants had to perform three tasks of finding: (1) the route by car from location A to location B, (2) a POI near location B, and (3) the car route from location A to location B, via location C.
- In the study by Khan and Adnan (2010), participants were asked to perform tasks: getting directions, printing a map, finding a location, and zooming and panning.
- Wang (2014), used the scenario where participants were travellers visiting for the first time the US capital, Washington, DC and they had to perform the following tasks:
    1. "Open" the service and choose the language of display of the information.
    2. Identify the White House (POI)
    3. Find additional information about it, such as a photo, its 3D model, etc.
    4. Locate a specific hotel in the area and find additional information about it, such as price and website.
    5. Choose the most suitable route from the POI to the hotel.

Overall, tasks should be limited in number and should reflect the context of using web mapping services (e.g., locating a POI). A scenario that defines the "role" of the participant when using the service (e.g., a visitor to a city) may help in making the survey process more attractive and enjoyable for participants.

Hence, all the tasks took place in the city of Boston, USA, giving participants the following scenario: "You have been admitted to MIT to attend some classes during a summer school. During your visit, you will get to know the city". Boston was chosen because as a US city, the mapping services are fully functional for that area and it fulfils the prerequisite of performing the tasks in an unfamiliar environment.

It was decided to test Google Maps first for all tasks for both groups since participants are by far more familiar with this web mapping service than any other tested in this study, serving the purpose of not discouraging participants in case they could not perform easily a task and of not putting too much effort to task execution at first attempt. Participants were told to use any functionality of the platform to perform the task(s). Survey administrators did not recommend any particular tool/method assessing how well participants could exploit or were familiar with the capabilities offered to them by each service.

For each task, the following data were recorded: (a) the successful/ unsuccessful execution of the task (score 1 or 0 respectively) and (b) the response time (time to complete the task regardless of success).

3.5.2 System Usability Scale

The second section includes the System Usability Scale (SUS) (Brooke, 1996). The SUS scale is a Likert psychometric scale, which consists of 10 closed-ended questions, with answers ranging from absolute disagreement (1) to absolute agreement (5). If the final rating is over 85 the system, service, or product is considered excellent to use. If it is between 70 and 85 acceptable to good, a rating of 50 to 70 indicates that it has usability issues and needs improvement, and finally, a system, service or product with a score less than 50 is considered unusable (Bangor et al., 2008).

The study used the Greek translation of SUS by Katsanos et al. (2012) who conducted surveys on the usability of learning management systems (LMS). The most important results of this research are: (a) the Greek version of SUS seems reliable (Cronbach's alpha=0.777), (b) the scale has good internal coherence and is one-dimensional.

4 Results

This section shows the results of each questionnaire, starting from the sample's digital profile and DNAS results (4.2), scores on STAT (4.3), performance on web-mapping services (4.4), and finally closing with correlations among these variables (4.5).

4.1 Data Pre-processing

All variables have been tested for normality with the Shapiro-Wilk test. Since the majority of variables do not follow normality assumptions, for most of the statistical tests, non-parametric measures have been applied unless otherwise indicated.

4.2 Participants' Profile and Digital Natives Self-Assessment

The reasons why participants use computers and smart devices are depicted on Fig. 1, with education, entertainment, and social networking being the three most prevalent.

Familiarity with each of the three web-mapping services is shown in Fig. 2, highlighting that participants are much less familiar with Here WeGo than Bing Maps, with Google Maps being the service participants are more familiar with, as expected.

Regarding the degree of familiarity with Boston, only 7% of the sample stated of having a moderate familiarity with Boston, satisfying the prerequisite that participants are not familiar with the area where the "scenario" took place.

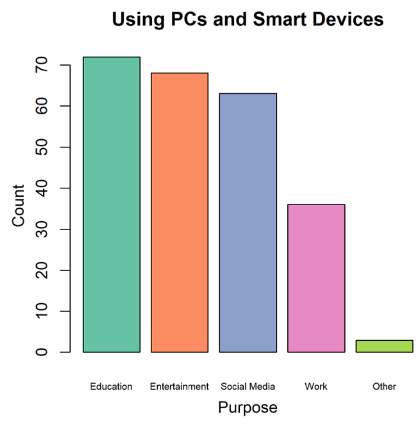

Figure 1. Reasons for using computers and smart devices

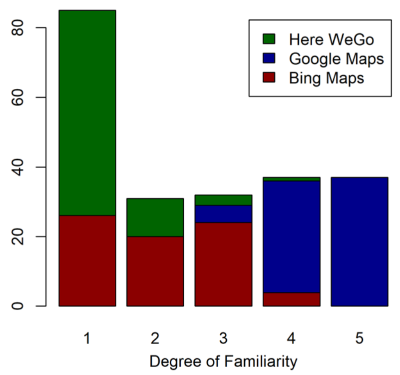

Figure 2. Google Maps is the most familiar web-mapping service among participants, with HERE WeGo being at the other end of the spectrum

Finally, DNAS presents good internal consistency (Cronbach's a = 0.802). The mean score of the sample is 105.51±13.98 (Fig. 3). As the score can range from 21 to 147, we can state that participants are self-assessed as moderate digital natives placing themselves above the midpoint of the score spectrum.

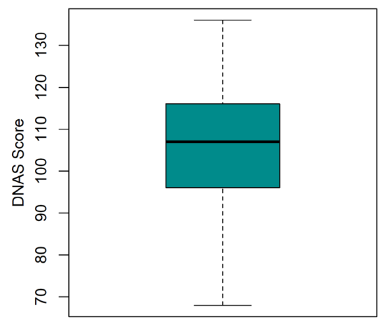

Figure 3. Boxplot of participants' DNAS scores

4.3 STAT Results

Internal consistency of STAT is poor (Cronbach's a = 0.549). However, accuracy of the participants' answers (Fig. 4) to each item in the Questionnaire are similar to previous studies (Lee and Bednarz, 2012); items 7 and 12 being problematic in almost each one of them. Thus, it was decided to leave the items intact and not to remove any of them from the final score which is the sum of all correct answers. Hence, our sample has a mean STAT score of 10.62±2.45, and no outliers are detected as confirmed by the boxplot (Fig. 5).

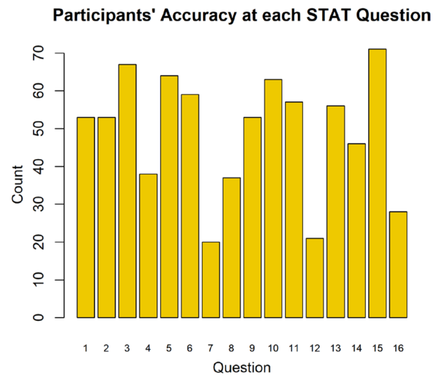

Figure 4. Number of participants correctly answering each STAT question

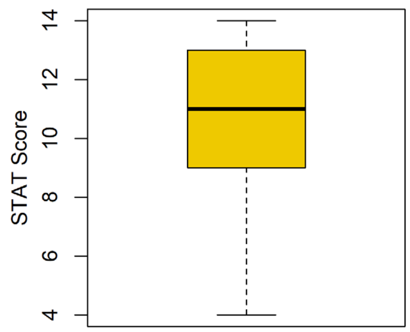

Figure 5. Boxplot of participants' STAT scores

4.4 Participants' Performance on Web Mapping Services

Figs. 6 and 7 show participants performance using the web-mapping services. Scores equal the sum of tasks accurately performed using each service; ranging from 0 to 5. It seems that participants performed better using Bing Maps than the other two services. In what follows, we present statistical tests results that support this assumption.

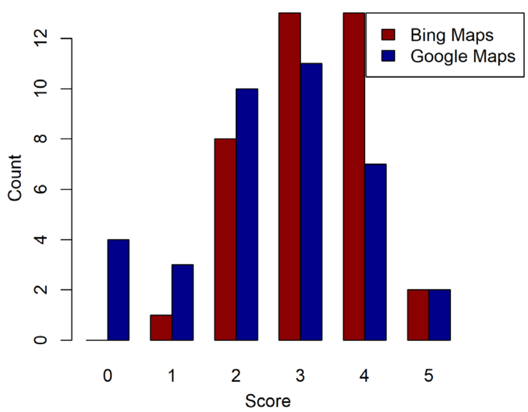

Figure 6. Tasks Scores of 37 Participants using Bing Maps; M=3.19 and Google Maps; M=2.54 (Group 1)

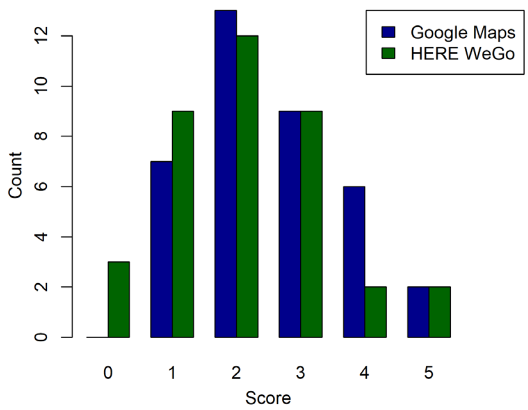

Figure 7. Tasks Scores of 37 Participants using Google Maps; M=2.54 and HERE WeGo; M=2.11 (Group 2)

4.4.1 Correctness in Task Performance

We compare task accuracy under two conditions (1) within groups; between GM and BM (Tab. 1)/ GM and HWG (Tab. 2) and (2) between groups; BM and HWG (Tab. 3).

Table 1. Differences in Success Rates between Google Maps and Bing Maps (Group 1)

Task	GM - BM	McNemar Test	
		X^2	p
1	**35.14% – 86.49%**	15.43	**<.001**
2	67.57% – 78.38%	1.50	.221
3	**56.76% – 21.62%**	7.58	**.006**
4	51.35% – 64.86%	1.45	.228
5	43.24% – 67.57%	2.78	.095

Table 2. Differences in Success Rates between Google Maps and HERE WeGo (Group 2)

Task	GM - HWG	McNemar Test	
		X^2	p
1	37.84% – 51.35%	0.84	.359
2	72.97% – 64.86%	0.57	.450
3	**62.16% – 29.73%**	7.56	**.006**
4	45.95% – 37.84%	0.80	.371
5	35.14% – 27.03%	0.27	.606

Table 3. Differences in Success Rates between Bing Maps and HERE WeGo

Task	BM - HWG	Fisher's Test	
		odds ratio	p
1	**86.49% – 51.35%**	5.91	**.002**
2	78.38% – 64.86%	1.95	.302
3	21.62% – 29.73%	0.66	.595
4	**64.86% – 37.84%**	2.99	**.036**
5	**67.57% – 27.03%**	5.48	**.001**

Results indicate better performance using Bing Maps over HERE WeGo in three tasks (Tasks 1, 4, and 5) and over Google Maps in one (Task 1), while participants are better using Google Maps when performing Task 3 over the other two services.

Comparing between overall scoring in task performance, we get the following results (Tab. 4) providing evidence that participants did better in executing the tasks using Bing Maps than using any of the other services.

Table 4. All tasks' scores comparisons show statistically significant difference between Bing Maps and Google Maps and HERE WeGo both. Between the last two no such difference is significant.

Tasks Scores Differences				
Within Groups	Wilcoxon matched-pairs signed-rank test			
	N	V	p	CI (95%)
GM - BM	37	78	**.012**	[-1.5, 0.0]
GM - HWG	37	281	.066	[0.0, 1.0]
Between Groups	Wilcoxon signed-rank test			
	N	W	p	CI (95%)
BM - HWG	37	1036	**<.001**	[1.0, 2.0]

4.4.2 Speed in Task Performance

We compare speed in task performance under two conditions (1) task successful completion time within groups; (a) in either of the two web mapping services (GM or BM/ GM or HWG), and (b) in both the web mapping services (GM and BM/ GM and HWG) and (2) task successful completion time between groups; in either of the two web mapping services (BM or HWG).

Since the majority of the tests indicated no significant differences in speed, results are omitted due to space limitations. Two statistically significant differences are calculated between participants' speed who performed Tasks 4 and 5 correctly using Bing Maps as opposed to those using HERE WeGo (Tab.5).

Table 5. Participants were faster using Bing Maps than HERE WeGo to successfully perform Tasks 4 and 5; (Task 4: $Md_{(BM)}$ = 84.5, $Md_{(HWG)}$ = 138.5, Task 5: $Md_{(BM)}$ =97, $Md_{(HWG)}$ = 213)

Between Groups	Wilcoxon signed-rank test			
BM-HWG	N/N	W	p	CI (95%)
Task 4	24/14	96.5	**.032**	[-96.0, -4.0]
Task 5	25/10	51.0	**.007**	[-164.0, -46.0]

From these results we cannot conclude whether participants were faster, without sacrificing accuracy, in any of the three web mapping services. Moreover, since all tasks needed the same strategy and the same number of steps in order to be executed successfully, regardless of mapping service used, we can assume that participants' speed is more or less the same using either service.

Since Google Maps was used first it the survey's task execution part, the repetition effect may have affected task accuracy and speed, favouring the service that was used second. However, no significant differences (improvements) in participants' accuracy (4.4.1) or speed (4.4.2) are achieved when performing the tasks on the second service. This is only holds for Task 1 in Group 1 where participants' success rate on Bing Maps is much better than on Google Maps (Tab 1).

4.4.3 SUS Results

Descriptive statistics and internal consistency of the SUS Questionnaire are shown on Tab 6 and 7, respectively.

Table 6. Mean, median and standard deviation of SUS Scores for each web mapping service

SUS Scores	Group 1		Group 2	
	GM	BM	GM	HWG
Mean	85.47	77.43	86.89	55.68
Median	85.00	77.50	87.50	57.50
SD	9.91	14.71	8.17	18.33

Table 7. Cronbach's alpha per group and service

Group	N	Service	Cronbach's alpha (α)
1	37	GM	0.778
		BM	0.869
2	37	GM	0.734
		HWG	0.910

For all services, Cronbach's alpha α>0.7 (Tab. 7), which is considered an acceptable value of internal reliability. Participants rated Google Maps higher than Bing Maps and HERE WeGo (Tab. 6), the statistical tests indicate that this difference is significant. The same holds between Bing Maps and HERE WeGo (Tab. 8).

Table 8. SUS scores differences between services

Differences in SUS scores				
Within Groups	Wilcoxon matched-pairs signed-rank test			
	N	V	p	CI (95%)
GM - BM	37	382.5	<.001	[5, 13.75]
	Paired t test			
	N	t	p	CI (95%)
GM - HWG	37	9.45	<.001	[24.5, 37.9]
Between Groups	Wilcoxon signed-rank test			
	N	W	sig.	CI (95%)
BM - HWG	37	1037.5	<.001	[12.5, 30.0]

4.5 Correlations among variables

According to the Fligner-Killeen test of homogeneity of variances, the two groups are homogeneous in terms of their scores in task performance for Google Maps (χ^2 = 3.15, df = 4, p-value = .533). Regarding SUS Scores for Google Maps, both groups exhibit similar behaviour as indicated by the Wilcoxon signed-rank test (W = 644.5, p-value = .668). Therefore, correlations can be calculated for Google Maps considering the sample as a whole (N=74). Tab 9 shows correlations between STAT, Familiarity with web mapping service, DNAS, and SUS and tasks score - TS for each web mapping service.

Table 9. Correlations between variables

STAT				
	N	z	Kendall's tau	p
TS GM	74	3.19	**0.29**	**.001**
TS BM	37	0.73	0.10	.462
TS HWG	37	1.67	0.22	.095
Web Mapping Service Familiarity				
	N	z	Kendall's tau	p
TS GM	74	0.86	0.09	.393
TS BM	37	2.07	**0.30**	**.038**
TS HWG	37	-1.43	-0.21	.151
DNAS				
	N	z	Kendall's tau	p
TS GM	74	0.66	0.06	.509
TS BM	37	1.91	0.25	.056
TS HWG	37	1.64	0.21	.101
SUS				
	N	z	Kendall's tau	p
TS GM	74	0.21	0.02	.831
TS BM	37	1.62	0.21	.105
TS HWG	37	-0.18	-0.02	.860

Results indicate that spatial ability as assessed by STAT is not correlated to user performance using web mapping services. Significant yet small positive correlation is calculated only between STAT scores and task accuracy using Google Maps.

DNAS scores do not correlate with accuracy on either web mapping service. DNAS has the shortcomings of all self-assessment instruments; it reflects self-imposed beliefs and conceptions that may not align with facts. Moreover, web mapping services may not be at the core of services and applications digital natives use and interact with as opposed to social media apps.

Regarding participants' familiarity with each web mapping service and their performance using them, correlations are insignificant too, with the exception of task accuracy using Bing Maps. In the case of familiarity, lack of correlations can be explained by the contradicting results; task accuracy using Google Maps was not as high as familiarity with it, task accuracy using Bing Maps was somewhat better than familiarity with it, and task accuracy using HERE WeGo was much better than its familiarity.

Regarding services evaluation through SUS, participants rated Google Maps higher than the other two. Nonetheless, participants were more accurate using Bing Maps than Google Maps, and no significant difference in accuracy has been identified between HERE WeGo and Google Maps (except for one Task).

4.6 Data and Software Availability

Statistical analyses have been performed in R (R Core Team, 2022) using: "tidyverse" (Wickham et al., 2019), "descTools" (Andri et al., 2022), "gdata" (Warnes et al., 2022), and "psych" (Revelle, 2022) packages. The collected survey data in tabular form and the R code of the statistical analysis workflow supporting this publication, are available on figshare and are accessible via the DOI: https://doi.org/10.6084/m9.figshare.22132940.v2.

The workflow underlying this paper was partially reproduced by an independent reviewer during the AGILE reproducibility review and a reproducibility report was published at https://doi.org/10.17605/osf.io/2em7v.

5 Discussion and Conclusions

Although this study' extra-laboratorial synchronous approach provided the opportunity of a large number of participants, bears uncertainty to some extent. The fact that administrators cannot control the environment where the study takes place, the condition of leaving participants unattended without knowing if they are distracted by other devices or services in their familiar environments or even the presence of other people in the same room are possible factors of biased results.

Although digital natives are familiar with the use of digital media, they are not as familiar with various functions of web mapping services, nor they seem flexible or adaptable enough to use successfully a service they are not that familiar with.

Overall, multiple factors affect web mapping performance: the service itself (interface and functionalities), poor knowledge of participants of the service, its interface and functionalities, and the complexity of the task to be performed, among others.

Finally, regarding spatial ability, it may be the case that successfully and quickly performing a task using a web mapping service is affected by other factors and may not be a spatial ability-related task but more of a digital task. Thus, future research may focus on determining how "spatial a task" is and how spatial ability relates to successfully and rapidly executing it when using a web mapping service.

Acknowledgments

The research work was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant" (Project Number: HFRI-FM17-2661).

Authors would like to thank all individuals who participated to the study.

References

Alghamdi, A.S., Al-Badi, A.L.I.H., Alroobaea, R., Mayhew, P. A: Comparative Study of Synchronous and Asynchronous Remote Usability Testing Methods. International Review of Basic and Applied Sciences 1, 61–97, 2013.

Andri S. et mult. al.: DescTools: Tools for Descriptive Statistics. R package version 0.99.45, https://cran.r-project.org/package=DescTools, 2022.

Bangor, A., Kortum, P.T., and Miller, J.T.: An Empirical Evaluation of the System Usability Scale. Int. J. Human-Computer Interaction 24, 574–594, doi:10.1080/10447310802205776, 2008.

Bednarz, R. and Lee J.: What improves spatial thinking? Evidence from the Spatial Thinking Abilities Test. International Research in Geographical and Environmental Education, 28:4, 262-280, DOI: 10.1080/10382046.2019.1626124, 2019.

Brooke, J.: SUS: A "Quick and Dirty" Usability Scale. In Usability Evaluation in Industry. CRC Press: London, ISBN 978-0-429-15701-1, 1996.

Enterprise Online Survey Software & Tools Available online: https://www.alchemer.com/ (accessed on 30 December 2022).

ISO 9241-11:2018: Ergonomics of Human-System Interaction — Part 11: Usability: Definitions and Concepts; ISO/TC 159/SC 4, 2018.

Çöltekin, A., Francelet. R., Richter, K-F, Thoresen, J. & Fabrikant S.I.: The effects of visual realism, spatial abilities, and competition on performance in map-based route learning in men, Cartography and Geographic Information Science, 45:4, 339-353, DOI: 10.1080/15230406.2017.1344569, 2018.

Katsanos, C., Tselios, N., and Xenos, M.: Perceived Usability Evaluation of Learning Management Systems: A First Step towards Standardization of the System Usability Scale in Greek. In Proceedings of the 2012 16th Panhellenic Conference on Informatics; pp. 302–307, 2012.

Kavouras, M., Kokla, M., Liarokapis, F., Pastra, K., and Tomai, E.: Comparative study of the interaction of digital natives with mainstream web mapping services. In Human-Computer Interaction. Design and User

Experience Case Studies: Thematic Area, HCI 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, Proceedings, Part III (pp. 337-350). Cham: Springer International Publishing, 2021.

Khan, Z.A. and Adnan, M.: Usability Evaluation of Web-Based GIS Applications: A Comparative Study of Google Maps and MapQuest, School of Computing, Blekinge Institute of Technology, 2010.

Lee, J., and Bednarz, R.: Components of Spatial Thinking: Evidence from a Spatial Thinking Ability Test. Journal of Geography, 111:1, 15-26, DOI: 10.1080/00221341.2011.583262, 2012.

Nivala, A.-M., Brewster, S., and Sarjakoski, T.L.: Usability Evaluation of Web Mapping Sites. Cartographic Journal, 45, 129–138, doi:10.1179/174327708X305120, 2008.

Prensky, M.: Digital Natives, Digital Immigrants Part 1. Horiz., 9, 1–6, doi:10.1108/10748120110424816, 2001.

R Core Team R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/, 2022.

Revelle, W.: psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois. R package version 2.2.9, https://CRAN.R-project.org/package=psych, 2022.

Sanchez C.A., and Branaghan, R.J.: The interaction of map resolution and spatial abilities on map learning. International Journal of Human-Computer Studies, 67(5), 475–481. https://doi.org/10.1016/j.ijhcs.2008.12.003, 2009.

Skarlatidou, A. and Haklay, M.: Public Web Mapping: Preliminary Usability Evaluation. GIS Research UK, 2006.

Teo, T.: An initial development and validation of a Digital Natives Assessment Scale (DNAS). Computer and Education 67, 51–57, https://doi.org/10.1016/j.compedu.2013.02.012, 2013.

Tullis, T., Fleischman, S., McNulty, M., Cianchette, C., and Bergel, M.: An Empirical Comparison of Lab and Remote Usability Testing of Web Sites. Orlando, FL, US, 2002.

Veenendaal, B., Brovelli, M.A., and Li, S.: Review of Web Mapping: Eras, Trends and Directions. ISPRS Int. J. Geo-Inf., 6, 317, doi:10.3390/ijgi6100317, 2017.

Wachowicz, M., Cui, L., Vullings, W., and Bulens, J.: The effects of web mapping applications on user satisfaction: an empirical study. In International Perspectives on Maps and the Internet, Peterson, M.P., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 397–415. ISBN 978-3-540-72028-7, 2008.

Wang, C.: Usability Evaluation of Public Web Mapping Sites. In Proceedings of the ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences; Copernicus GmbH, Vol. XL–4, 285–289, 2014.

Warnes, G. R. et mult. Al.: 'gdata': Various R Programming Tools for Data Manipulation Version 2.18.0.1, https://CRAN.R-project.org/package=gdata, 2022.

Wickham, H.et mult. Al.: Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686. doi:10.21105/joss.01686https://doi.org/10.21105/joss. 01686, 2019.