Received: 6 December 2023 | Revised: 20 June 2024 | Accepted: 27 June 2024

DOI: 10.1111/ssm.18303

RESEARCH ARTICLE

WILEY SSMA

Design and usability evaluation of a mobile app for elementary school inquiry-based science learning

Iraya Yánez-Pérez | **Radu Bogdan Toma** | **Jesús Ángel Meneses-Villagrá**

Department of Specific Didactics, Area of Science Education, University of Burgos, Burgos, Spain

Correspondence
Iraya Yánez-Pérez, Department of Specific Didactics, Area of Science Education, University of Burgos, Burgos, Spain.
Email: iyanez@ubu.es

Funding information
Agencia Estatal de Investigación, Grant/Award Number: PID2020-117348RB-I00 - 10.13039/501100011033

Abstract

Teachers often struggle to implement inquiry-based science teaching. To support them, IndagApp—a 3D educational app that offers curriculum-aligned, inquiry-based lesson plans—was designed. The app is rooted in the inquiry phases recommended in best-practices literature, which arguably align with most international standards. This study describes IndagApp and evaluates its usability with fifth-graders from two elementary schools in Spain ($N = 43$). The system usability scale (SUS) and the Pictorial-SUS were used to collect students' feedback on usability. Both frequentist and Bayesian analyses were conducted to compare the mean SUS score with established benchmarks for usability. The results showed that IndagApp had high usability ratings, with most students rating it as the "Best imaginable" or "Good." The mean SUS score was 84.816, which was significantly higher than the benchmark score of 68. There were no differences in usability between girls and boys, and students in private and public schools. These findings suggest that IndagApp is a valuable resource for inquiry learning in elementary grades and has significant implications for science education and science teacher professional development, as it helps teachers adopt reform-oriented teaching practices that align with the curricular standards and goals.

KEYWORDS
elementary education, ICT resource, IndagApp, inquiry, usability

1 | INTRODUCTION

Inquiry-based science teaching is a student-centered approach that fosters hands-on, constructivist-oriented learning experiences (de Jong et al., 2023; Romero-Ariza et al., 2019; Toma, 2022a). This pedagogical approach has become pivotal in elementary science education. It engages students in exploring scientific concepts through the adoption of scientific practices (Crawford, 2014; García-Carmona, 2020). These include asking questions, designing investigations, collecting and analyzing data, and generating evidence-based arguments, among others (NGSS Lead States, 2013; Osborne, 2014). Through inquiry units, students develop a deeper understanding of scientific principles and also cultivate essential 21st-century skills, like problem-solving, teamwork, and improved attitudes toward science (Liou, 2021; Ma, 2023).

The benefits of inquiry-based science teaching are well-supported by research. Literature reviews and meta-analyses report benefits regarding improved attitudes toward science (Aguilera & Perales-Palacios, 2020) and performance success (Lazonder & Harmsen, 2016). Hence, curricula reforms worldwide have stressed the importance of enacting inquiry teaching for decades (Abd-El-Khalick et al., 2004; LOMLOE, 2020; NGSS Lead States, 2013).

2 | CHALLENGES IN ENACTING INQUIRY

Despite the advantages of inquiry teaching and learning, there remain challenges in effectively implementing inquiry-based approaches (Romero-Ariza et al., 2019). Teachers face resource constraints and time limitations (Baroudi & Helder, 2019; Chichekian et al., 2016; Krämer et al., 2015). Developing and delivering inquiry-based lesson plans need adequate materials and careful planning. Striking a balance between curriculum coverage and facilitating student-centered inquiry practices is also demanding. Another notable difficulty lies in understanding what is inquiry teaching (Toma et al., 2017). The varying interpretations of inquiry-based methods (Rönnebeck et al., 2016) make it a challenging endeavor for teachers, which usually confuses it with hands-on, recipe-like laboratory experiments. Hence, teachers tend to adopt approaches that fall short of inquiry principles (Cañal et al., 2016; García-Carmona et al., 2018; Romero-Ariza et al., 2019).

3 | INFORMATION-COMMUNICATION-TECHNOLOGY RESOURCES FOR INQUIRY TEACHING

To address challenges in inquiry teaching, teachers need explicitly designed resources to support inquiry methodologies. In the past decade, a surge in information-communication-technology (ICT) resources has occurred (Oliveira et al., 2019; Raman et al., 2022). Such resources include educational apps, lab simulations, and interactive learning platforms. Noteworthy initiatives have been undertaken, such as PhET Colorado (Wieman et al., 2008) and ChemCollective (Yaron et al., 2010). Their use has exhibited favorable outcomes for learning (Scalise et al., 2011), improving affective domains like attitudes and motivation (Marrero-Galván & Hernández-Padrón, 2022), and teaching students self-regulation strategies (Reginald, 2023). While these resources represent high-quality ICT, many resemble recipe-like laboratory practices, where students use simulation software to only confirm specific phenomena, and most of them present too much information, which may cause cognitive load to students (Ali et al., 2022). Their focus, therefore, is on the experimental simulation, lacking contextualization and neglecting important inquiry phases such as the orientation or conceptualization phase (Pedaste et al., 2015).

Indeed, existing ICT resources are limited in supporting inquiry-based learning. In their literature review of 79 studies, Scalise et al. (2011) identified a proliferation of ICT resources, such as virtual labs or simulations. However, only 17% of such resources were designed to explicitly address inquiry-related practices. For example, existing resources, such as PhET Colorado, are not rooted in a real-world problem and there is no explicit phase for research question formulation or hypothesis development. Furthermore, they do not allow for explicitly identifying dependent, independent, and control variables; this aspect may limit students' ability to design controlled investigations, which is an important inquiry skill to be developed (de Jong et al., 2023; García-Carmona, 2020). Likewise, data presentation is another shortfall; given absent tables and graphs, students miss opportunities to develop data analysis and interpretation skills. Finally, existing simulations lack guidance for students and explanations of the scientific phenomena represented, which may leave the learning experience incomplete. On the other hand, systematic literature reviews by Potkonjak et al. (2016) and Reeves and Crippen (2021) revealed that the current use of virtual labs and simulations lacks a theoretical foundation, and the ICT tools are often confused with teaching methods instead of educational resources.

4 | THE PRESENT STUDY

Inquiry-based science teaching is strongly advocated by educational curricula and research. However, teachers face numerous obstacles in implementing this method effectively. Despite many ICT resources available, there is a significant shortage of tools that fully support inquiry teaching. Against this background, the authors of this study designed "IndagApp" (in Spanish, Inquiry-App), an innovative 3D educational app tailored for inquiry-based science teaching in elementary grades (Yánez-Pérez et al., 2024a). IndagApp has been designed to overcome the limitations of the existing simulations mentioned above. Hence, it offers curriculum-aligned lesson plans rooted in the inquiry phases recommended by Pedaste et al. (2015), which resemble the vision of inquiry in most standards, such as the Next Generation Science Standards in the USA (NGSS Lead States, 2013), the MINEDUC (2012) curriculum in Chile or the LOMLOE (2020) educational law in Spain. The goal of IndagApp is to assist science teachers in using inquiry. Previous research with pre- and in-service teachers reported high levels of usability (Yánez-Pérez et al., 2024a, 2024b). The results show that IndagApp has a high level of technological and pedagogical usability. It was rated as easy to use, with good esthetics and system efficiency, and provides support and guidance for inquiry-based teaching. Both pre-service and in-service teachers appreciated its intuitive design, dynamic visual learning experience, and effective integration of user interface functions.

The importance of usability testing for educational resources cannot be overstated. The International Organization for Standardization (ISO, 2018) defined usability 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). Prior literature reviews indicate a focus on usability evaluations in higher education, with varying user perceptions based on the type of educational technology (Lewis, 2018; Vlachogianni & Tselios, 2022). Mobile applications and multimedia resources tend to achieve satisfactory usability levels compared to existing benchmarks.

The aim of this study, therefore, was to evaluate the usability of IndagApp with end users, elementary school students. This endeavor was necessary to ensure that the resource could be used by both girls and boys from public and private schools, with the latter group being likely to have more access to ICT resources. By doing so, this research contributes valuable insights into the feasibility of using IndagApp for inquiry-based science teaching. Usability testing is essential for educational materials, ensuring they meet the needs of teachers and students effectively. Lewis (2014, 2018) emphasizes its role in the success of ICT tools. Without proper usability, assessing learning outcomes, motivation, and attitudes is challenging and untrustworthiness. The research questions are three-fold:

RQ1. What is the perceived usability of IndagApp among elementary school students?

RQ2. To what extent does gender influence the perceived usability of IndagApp?

RQ3. To what extent do the types of schools (public vs. private) influence the perceived usability of IndagApp?

5 | THE EDUCATIONAL RESOURCE: IndagApp

5.1 | General characteristics

IndagApp is a 3D educational app designed for inquiry-based science education for upper elementary/middle grades, with students aged 10–14. It supports Android ≥5.0 smartphones or tablets and PCs with Windows 7+ or above software. The app is freely accessible in Spanish on Google Play (https://play.google.com/store/apps/details?id=com.ITACA.Indagapp&hl=es_419) for Android and PC access can be requested from the corresponding author or the project website (https://www.webciencia.es/index.php/ind-virtual). IndagApp comprises 10 inquiry units about different phenomena: plant growth, crystal formation, forces, flooding, bacterial growth, photosynthesis, buoyancy, valley formation, light refraction, and balloon flight. As can be noted, the units are related to physics, chemistry, biology, and geology school subjects or curricula. The interface was adapted to the preferences of students aged 10–14 y/o, featuring characters and graphics that resemble the video games that are commonly played by this age group; for example, Figure 1 shows the main characters of an inquiry unit about balloon flight, with a style similar to the games Roblox or Minecraft.

To facilitate the use of IndagApp in the classroom, and because of previous usability testing studies with in-service and pre-service teachers (Yánez-Pérez et al., 2024a, 2024b), support resources are provided in the form of student workbooks (Figure 2). These workbooks are available for free download from the project

FIGURE 1 Example of the graphics for the balloon flight inquiry unit.

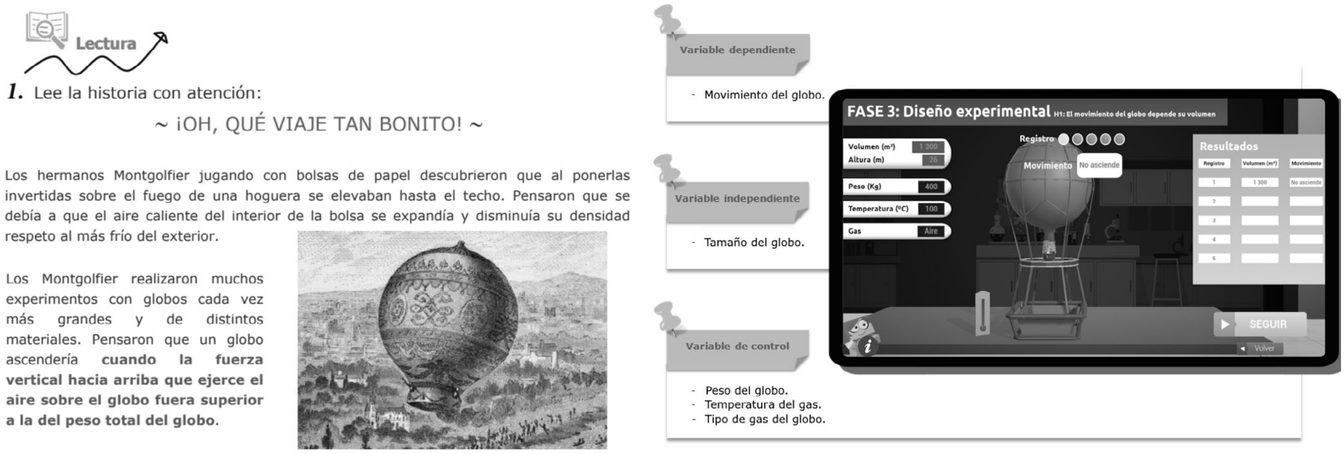

FIGURE 2 Example of workbooks.

website (anonymized link) and are compatible with PCs or Tablets as fillable PDF files, which minimizes the environmental impact. Alternatively, they can also be printed out if needed. At the moment of the publication of this manuscript, the app is in the process of being translated into English and Portuguese to enable its validation and use in other educational contexts and to support its international implementation.

5.2 | Content and inquiry phases

Each inquiry unit is rooted in a standardized structure, adapted from Pedaste et al. (2015) literature review (Figure 3). Some phases were renamed after pilot studies (Yánez-Pérez et al., 2024b) to improve the comprehensibility of the inquiry process.

The orientation phase, referred to as the "Statement of the Problem" engages students by presenting a story of an everyday life experience (Figure 4). For example, the inquiry regarding plant growth begins with a story about plants and their vital needs. Next, during the conceptualization phase, now called "Research question and hypothesis" students are encouraged to ask research questions about the problem introduced. The aim is to ask scientific research questions that could be addressed through an experimental design. Then, students are presented with the research question and the four potential hypotheses, facilitating the process of hypothesis formulation.

The investigation phase of Pedaste et al. (2015) consists of two phases in IndagApp: "Experimental design" and "Results and interpretation." In the experimental design phase, students identify dependent, independent, and control variables to test the proposed hypotheses, hence learning about experiment design. In the results and interpretation phase, students test each hypothesis and interpret findings. To do so, virtual simulations were

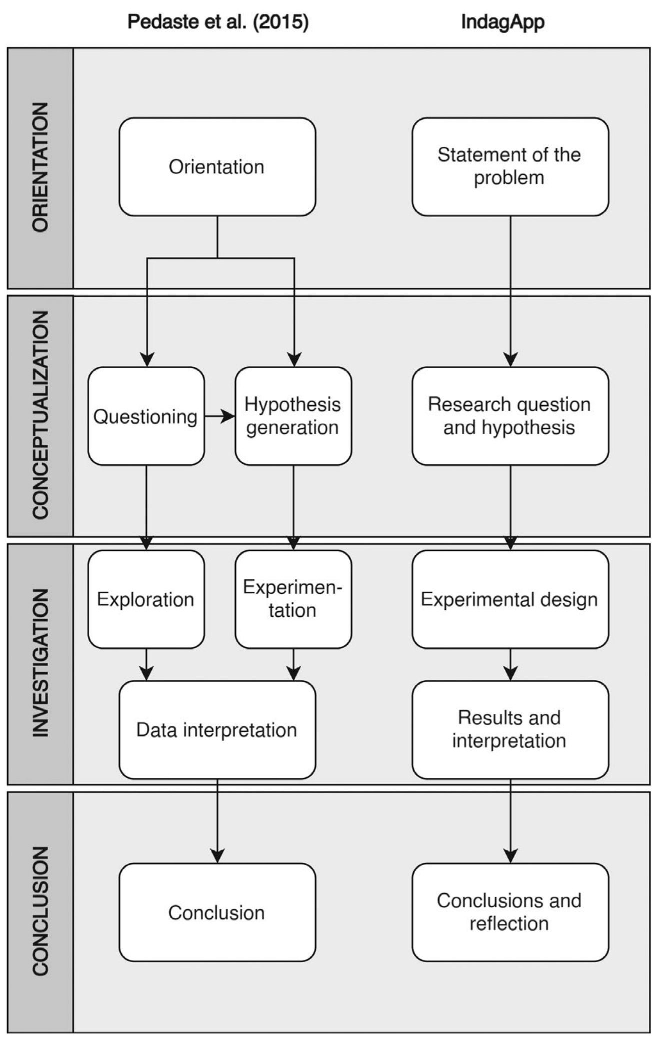

FIGURE 3 Adaptation of Pedaste et al. (2015) inquiry phases for IndagApp.

specially designed for IndagApp (Figure 5), with careful consideration to avoid excessive information or seductive details (e.g., flashy animations) that may produce

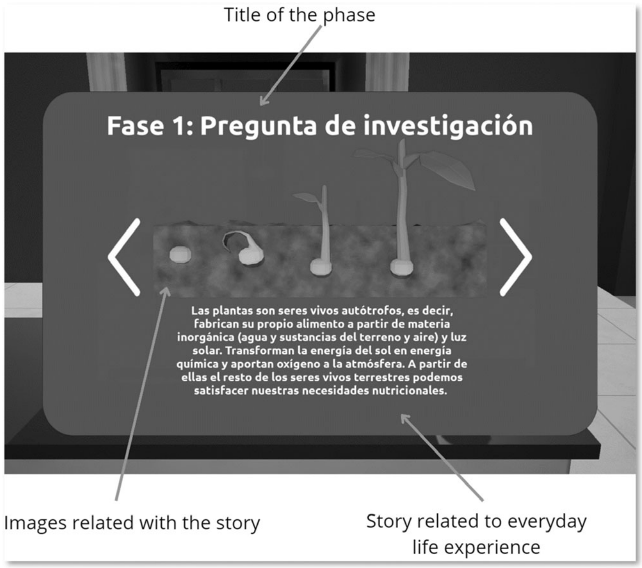

FIGURE 4 Orientation phase screen layout.

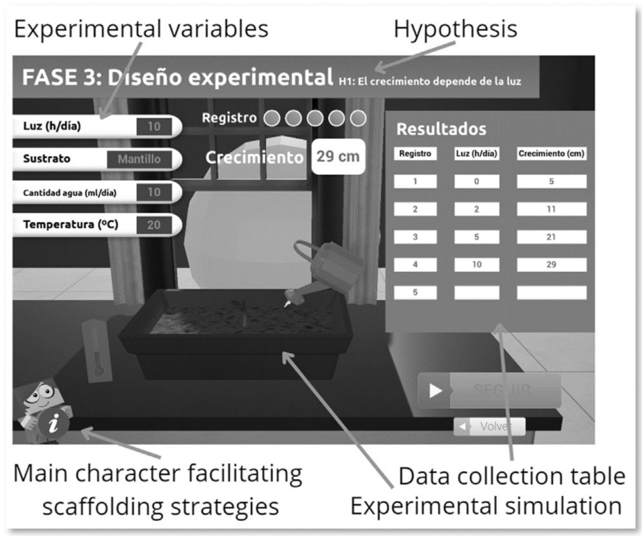

FIGURE 5 Screen layout for the experimental simulation.

cognitive load (Sundararajan & Adesope, 2020). Students gather five data sets by manipulating and controlling experimental variables. The collected data are organized into a table and figure for visual interpretation. Students assess their correct interpretation of the results by responding to questions related to the data.

Lastly, the conclusion phase, renamed "Conclusions and reflection" assesses students' understanding. To do so, reinforcement and application questions are used. Essentially, this phase evaluates students' understanding of the phenomenon investigated and the subject matter content, as well as their ability to apply acquired knowledge to other real-life contexts.

5.3 | Scaffolding strategies and guidance

IndagApp includes scaffolding strategies in video or text form (Zacharia et al., 2015). They help students in different phases of the inquiry process. The main character of the story guides the student along the way, giving instructions, advice, and short videos on various topics; for example, explaining the dependent, independent, and control variables. On the other hand, IndagApp comprises both guided and structured inquiry units. The inquiry learning process varies depending on the amount of teacher guidance (Vorholzer & von Aufschnaiter, 2019). In short, during a confirmation-type inquiry lesson, students confirm a phenomenon by following instructions and knowing the research question, procedure, and results beforehand. In structured inquiry, students receive the research question and the experimental procedure, but not the results. In guided inquiry, students are given the research question, but they design the experimental procedure and do not know the results in advance. Finally, in open inquiry, students generate the research question, design the procedure, and gather results with minimal teacher guidance. Regarding IndagApp, all units include the research question and hypothesis to be tested. However, some units also provide the experimental design conditions for the control variables (hence, being a structured inquiry unit) while others leave it open to students' decisions (hence, resembling a guided inquiry unit). This is done to facilitate student understanding and teacher implementation of inquiry (Toma, 2022b; Zacharia et al., 2015). Therefore, IndagApp aligns with recent evidence on the significance of scaffolding for inquiry teaching and learning (de Jong et al., 2023; Strat et al., 2023).

6 | METHODS

6.1 | Study design

This study is part of a design-based research (DBR) project. DBR aims to develop and refine interventions or products collaboratively with practitioners (Plomp, 2013). Using DBR, this project tackles a real-world problem in authentic learning environments: the gap in ICT resources explicitly addressing all relevant phases of inquiry-based science teaching (Pedaste et al., 2015). The process involved iterative phases of design, implementation, evaluation, and refinement of the intended product, IndagApp (Pool & Laubscher, 2016). Within the DBR framework, in previous studies, we used empirical evidence to continuously improve IndagApp, creating an effective educational innovation.

During the first phase of the project, named preliminary research, teachers' needs for implementing inquiry were identified, context analysis was performed, and existing inquiry literature was synthesized to provide best-practice guidelines. Based on this information, a series of app mock-ups were developed and refined with feedback from university experts in science education, inclusive education, and computer engineering. This resulted in the design of the first version of IndagApp. In the second phase of the project, named prototyping or formative evaluation, the initial version of the app was tested iteratively, improving its design and content with each new iteration.

Specifically, the app underwent two rounds of formative evaluation before this study. First, a panel of science education experts and in-service teachers reviewed it and gave feedback for improvement (Yánez-Pérez et al., 2024a). Then, pre-service teachers tested it and recommended minor changes (Yánez-Pérez et al., 2024b). These led to the potential third and final version of IndagApp that this study evaluates. The design of the app will be deemed finalized, thus constituting the last and final version, if the usability testing conducted in this study with elementary school students yields satisfactory results against established benchmarks for the system usability scale (SUS), as explained below (Lewis, 2018). After that, the third and final phase of the project, the summative evaluation, will examine the impact of using IndagApp on outcomes of interest, such as attitudes, achievement motivations, and development of inquiry procedural skills. Ethical approval was secured from the bioethics committee at the University of Burgos.

6.2 | Participants

This study used convenience sampling to recruit participants from two elementary schools in (anonymized), Spain, one public ($n = 20$) and one private ($n = 23$). The schools are located in Castile and León; hence, both public and private schools were included in the study to represent its educational milieu. It should also be noted that this region is the one with the highest achievement in science and mathematics in Spain. According to the 2019 edition of the Trends in International Mathematics and Science Study (TIMSS2019), the region scored 535 and 528, respectively, while the Spanish average was 511 and 502 (526 and 527 for Organisation for Economic Co-operation and Development countries). The participants were fifth-graders. There were 18 girls (41.9%) and 24 boys (55.8%); one student's gender was unknown. The ages were 9 (14%), 10 (72.1%), and 11 (11.6%) years.

6.3 | Instruments

The SUS, which comprises 10 Likert-type items, was used for data collection (Brooke, 1996). The SUS is a gold standard instrument for measuring ICT usability (Lewis, 2018). SUS provides user feedback on usability through an overall score representing the unidimensional construct. To enhance students' comprehension, we complemented the written SUS with the Pictorial-SUS (Baumgartner et al., 2021), which adds a pictorial representation of the written items (Figure 6). By doing so, usability assessment is more accessible and user-friendly, especially for students in primary education. Examples of items include “I thought IndagApp was easy to use” and “I felt very confident using IndagApp.”

Cronbach's alpha ($\alpha = 0.716$) and McDonald's omega coefficient ($\omega = 0.673$), which is a better indicator of reliability for Likert-type items, revealed adequate internal consistency reliability for preliminary research (Hayes & Coutts, 2020) and in alignment with existing research (Lewis, 2018).

6.4 | Procedure

The usability of IndagApp was evaluated in a real classroom environment. Students were organized into pairs. Each pair was given a 10′ Tablet with the app installed. Students used the IndagApp resource as intended in its design. That is, following the teachers' instructions, in line with the principles of guided inquiry (Vorholzer & von Aufschnaiter, 2019). Two different inquiry units, about forces and plant growth, were implemented in separate sessions to ensure extensive use of the app. These units were selected to engage both boys and girls in science learning, based on previous studies that indicated a gender difference in content preferences, with elementary school boys favoring physics topics, while girls tend to favor biology topics (Toma, 2022c). Each implementation spanned three sessions, with each session lasting 1 h. After completing the units, the students responded to the paper-and-pencil usability instrument anonymously and confidentially.

6.5 | Data analysis

The SUS scoring method was used (Brooke, 2013). It consists of subtracting 1 from odd-numbered items and subtracting the participant's score from 5 for even-numbered items; then, the resulting score is summed and multiplied by 2.5 to obtain a scale ranging from 0 to 100. Scores equal to or greater than 68 indicate adequate usability.

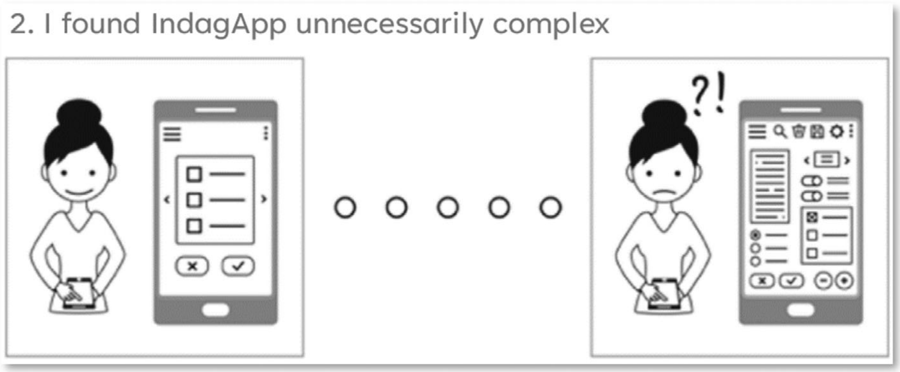

FIGURE 6 Format of the Pictorial-SUS questionnaire used.

Bangor et al.'s (2009) adjective scale was next used for a qualitative interpretation of the findings, following recommended descriptors and cutoff values (Lewis, 2018): worst imaginable (12.5), awful (20.3), poor (35.7), fair (50.9), good (71.4), excellent (85.5), and best imaginable (90.9).

To examine whether the sample mean significantly differed from the cutoff value of 68, a frequentist and Bayesian one-sample Wilcoxon signed rank test was conducted since the Shapiro–Wilk test suggests a deviation from normality ($W = 0.654$, $p < 0.001$). Furthermore, the SUS questionnaire is scored on a five-point ordinal Likert scale. Consequently, nonparametric tests are more appropriate than parametric tests because the assumption of a continuous dependent variable does not hold (Knapp, 2018).

To examine differences based on gender (girls and boys) and school type (public and private), frequentist and Bayesian Mann–Whitney U tests were conducted. Frequentist and Bayesian analyses are two distinct approaches with different assumptions and interpretations. Conducting both types of analyses offers advantages: (i) cross-validating results to highlight inconsistencies or false positives (type I errors), (ii) providing a comprehensive and more nuanced understanding of the phenomenon studied, and (iii) ensuring robustness by confirming results across analyses (for more rationale, see Ioannidis, 2019; Ly et al., 2020).

6.6 | Sample size and power analysis

The study adhered to sample size recommendations for quantitative measurements in usability testing. SUS studies require usually 20–30 users for sufficient precision, confidence, and variability; some previous SUS research achieved 100% agreement with 12 users (Lewis, 2018). A power analysis with G*Power software was conducted to ensure an adequate sample size. A one-sample Wilcoxon signed rank test with 80% power and medium effect size required at least 35 participants. Based on previous research using the SUS (Lewis, 2018), and to reduce the likelihood of type I error (false positive), a medium effect size was used. Thus, this study, with $N = 43$, had enough statistical power to assess the usability of IndagApp against cutoff criteria (Perugini et al., 2018).

7 | RESULTS

7.1 | Usability score

The SUS average score was 84.816 (SD = 14.728). The need for assistance ("I think that I would need the support of my teacher to be able to use IndagApp," M = 3.24, SD = 1.50) and frequency of use ("I think that I would like to use IndagApp frequently," M = 3.79, SD = 1.23) were the items with more negative ratings. However, both scores are considered positive, considering a response scale of five options. Conversely, ease of use ("I thought IndagApp was easy to use," M = 4.53, SD = 0.77) and integration of functionalities ("I found the various functions in IndagApp to be were well integrated," M = 4.70, SD = 0.56) received the highest ratings.

A qualitative interpretation using Bangor et al.'s (2009) adjective scale suggested that the usability of IndagApp was the best imaginable for 67.44% of the students and good for 20.93%. Only 9.3% rated it as fair and one student as poor. None of them considered it as the worst imaginable or awful.

A frequentist one-sample Wilcoxon signed-rank test indicated that the mean score was statistically significantly higher than the SUS cutoff score of 68 for adequate usability, $Z = 110.272$, $p < 0.001$. Similarly, Bayesian analysis yielded a high Bayes factor ($BF_{10} = 370.653$, $W = 857$), indicating strong evidence for the alternative hypothesis (H1), which suggests that students' SUS scores are much higher than the cutoff score (van Doorn et al., 2021). The prior-posterior plot (Figure 7) shows that with 95% probability, the true effect size (δ) is in the range of 0.549 and 1.337, and the posterior median is 0.945, indicating a large effect (Ly et al., 2020).

7.2 | Gender and school comparisons

Frequentist Mann–Whitney U test indicated no statistically significant difference between girls' and boys' SUS scores, $W = 248.500$, $p = 0.329$ (Figure 8a), nor between students in public and private schools, $W = 166$, $p = 0.061$ (Figure 8b).

Bayesian analysis supported these findings. For gender (Figure 9a), $BF_{10} = 0.398$ and δ within 0.447 and 0.819 indicates evidence for the null hypothesis (H0 = no gender differences). For school type (Figure 9b), $BF_{10} = 0.794$ and δ within −1.056 and 0.209 also indicate evidence for the null hypothesis of no school-type differences. Taken together, these findings suggest that IndagApp's usability is consistently high across girls and boys in public and private schools.

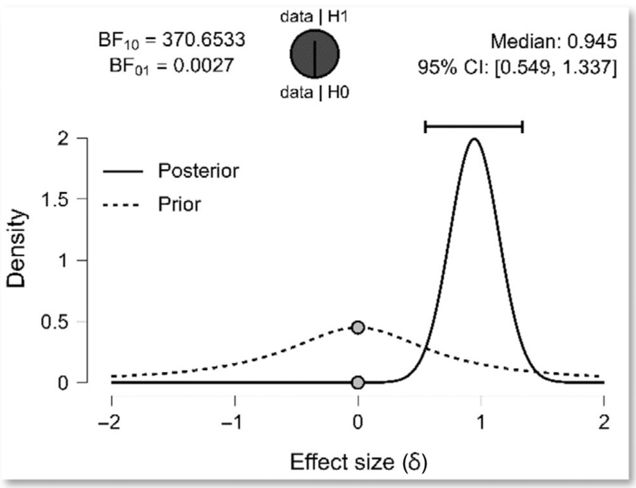

FIGURE 7 Bayesian one-sample Wilcoxon signed-rank test.

8 | DISCUSSION

This study focused on the usability evaluation of the educational application named IndagApp. The app was designed to address limitations in existing ICT resources, such as the lack of adherence to all phases of scientific inquiry or limitations regarding design features causing cognitive load (Ali et al., 2022; Scalise et al., 2011). IndagApp was specifically designed to facilitate inquiry-based science instruction within elementary grade levels by rooting the inquiry units into the steps of the inquiry cycle reported by Pedaste et al. (2015). After conducting a usability evaluation with elementary school students, the results indicate a notably high usability against standardized benchmarks (Lewis, 2018). Scrutiny of the scores revealed that most students appraised the app as the "Best Imaginable" and that no gender and school type (private or public) differences in usability were identified.

Taken together, these findings suggest that IndagApp exhibits a high level of usability among students aged 9–12. This usability evaluation is based on previous research conducted with science education experts, pre-service, and in-service teachers (Yánez-Pérez et al., 2024a, 2024b). In those studies, numerous improvements were made, including the correction of errors, the redesign of the interface to suit students' interests and age, the inclusion of scaffolding using written messages and videos, and the adaptation of the inquiry phases by using vocabulary that is more accessible. Based on the findings of this study, it seems that these improvements

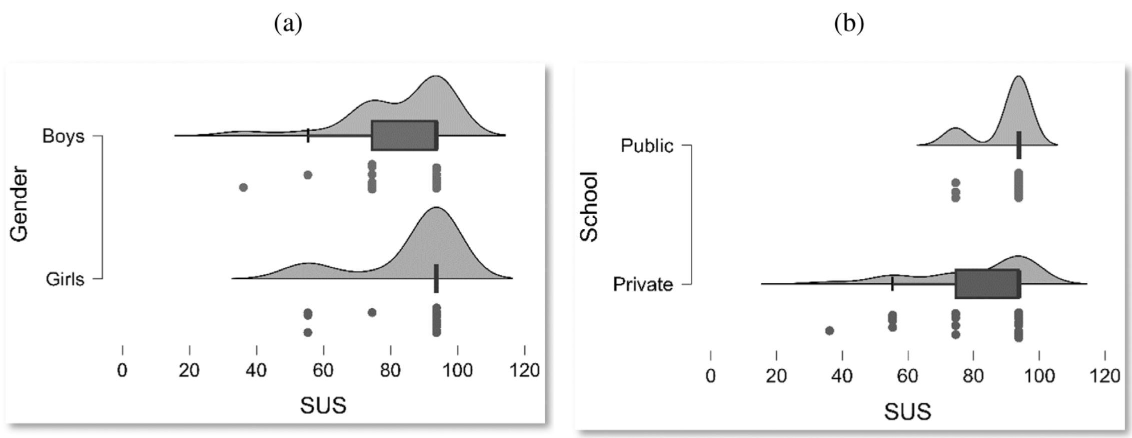

FIGURE 8 Frequentist Mann–Whitney U tests. (a) Gender. (b) School.

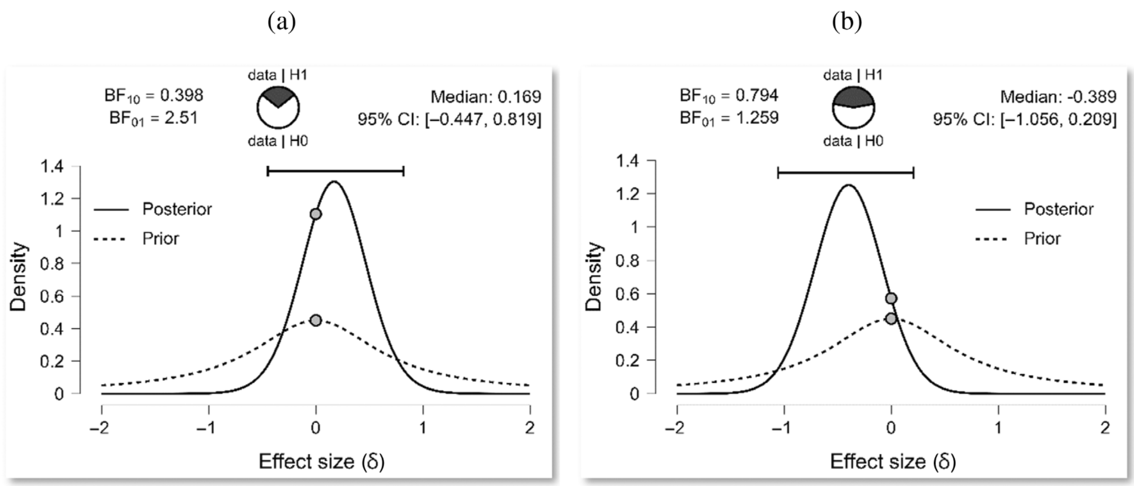

FIGURE 9 Bayesian Mann–Whitney U tests. (a) Gender. (b) School.

were effective in making IndagApp highly usable for elementary school students.

This study, therefore, underscores the potential of IndagApp as an effective tool to support inquiry-based science education. Teachers cite a lack of preparation time and instructional resources as major obstacles to implementing inquiry approaches (Baroudi & Helder, 2019). Thus, access to ready-to-use inquiry resources is important, especially in the Spanish context where teachers rely heavily on textbooks (Romero-Ariza et al., 2019) and hold conceptions aligned with traditional, lecture-based teaching approaches (García-Carmona et al., 2018).

8.1 | Implications

The results have significant implications for both teachers and students. For in-service teachers, IndagApp can ease the adoption of inquiry-based science teaching. It contains curriculum-aligned lesson plans and a standardized structure rooted in extant literature (Pedaste et al., 2015). It also includes ready-to-use students' workbooks and scaffolding guidelines (Zacharia et al., 2015). Thus, it may help teachers use inquiry-driven lessons across various phenomena related to physics, chemistry, biology, and geology. Ultimately, this may align their teaching with reform-oriented practices and pedagogies (Fang, 2020; Romero-Ariza et al., 2019).

For pre-service teachers, IndagApp offers a valuable tool to learn about inquiry teaching methodology. Incorporating IndagApp into training plans may help them gain practical experience in such an approach and benefit their professional development. Extant literature suggests that prospective teachers need to engage in inquiry lessons to be able to implement them (Capps et al., 2012; Toma et al., 2017; Zeivots et al., 2023).

Implications for students should also be taken into consideration. Since students rated the app highly usable, IndagApp may enhance their science learning experiences and encourage active participation. Indeed, reviews of the literature conclude that the use of ICT resources has been linked to improved learning-related outcomes (Scalise et al., 2011; Vlachopoulos & Makri, 2017) and attitudes (Lee & Tsai, 2013; Reeves & Crippen, 2021; Rutten et al., 2012). Therefore, IndagApp has the potential to positively influence these outcomes.

8.2 | Limitations

A few limitations should be considered. First, this study only evaluated the app's usability, not its effects on students' learning or other outcomes. This is because this investigation is part of an ongoing DBR project that aims to establish the app's usability before large-scale implementations (Plomp, 2013). Consistent with the DBR methodology (Hoadley & Campos, 2022), future research will investigate the app's long-term impacts. Second, this study was conducted with a convenience sample, which limits the generalizability of the findings. Third, the usability evaluation was performed with students from a specific Spanish educational region, characterized by high TIMSS results in science and mathematics. The results, therefore, may vary across different regions and student populations with lower educational achievements. A more diverse sample would enhance the external validity of the conclusions. Finally, the authors of IndagApp acknowledge that the app does not represent the full scope and conceptualization of scientific inquiry

(Crawford, 2014; Rönnebeck et al., 2016), but rather emphasizes investigations that require an experimental design. For example, IndagApp does not address important aspects of scientific practices, such as modeling. Yet, it should be noted that this decision to focus on experimental types of investigations is consistent with curricular reforms that highlight the importance of designing and conducting experimental investigations in science education (LOMLOE, 2020; NGSS Lead States, 2013).

Despite these limitations, this study provides evidence for the usability of IndagApp, highlighting its potential for inquiry-based science learning in elementary grades. The results regarding its usability, combined with its alignments with best practices in inquiry teaching, make it a valuable resource. Further research is being planned to gain a comprehensive understanding of its impact on student conceptual understanding, development of inquiry-procedural skills, and improvement of affective variables such as motivation and attitudes.

ORCID

Iraya Yánez-Pérez https://orcid.org/0000-0002-5260-2228
Radu Bogdan Toma https://orcid.org/0000-0003-4846-7323
Jesús Ángel Meneses-Villagrá https://orcid.org/0000-0003-4839-0418

REFERENCES

Abd-El-Khalick, F., Boujaoude, S., Duschl, R. A., Lederman, N. G., Mamlok-Naaman, R., Hofstein, A., Niaz, M., Treagust, D., & Tuan, H. (2004). Inquiry in science education: International perspectives. *Science Education*, *88*(3), 397–419.

Aguilera, D., & Perales-Palacios, F. J. (2020). What effects do didactic interventions have on students' attitudes towards science? A meta-analysis. *Research in Science Education*, *50*(2), 573–597. https://doi.org/10.1007/s11165-018-9702-2

Ali, N., Ullah, S., & Khan, D. (2022). Interactive laboratories for science education: A subjective study and systematic literature review. *Multimodal Technologies and Interaction*, *6*(10), 1–34. https://doi.org/10.3390/mti6100085

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

Baroudi, S., & Helder, M. R. (2019). Behind the scenes: Teachers' perspectives on factors affecting the implementation of inquiry-based science instruction. *Research in Science and Technological Education*, *1–22*, 68–89. https://doi.org/10.1080/02635143.2019.1651259

Baumgartner, J., Ruettgers, N., Hasler, A., Sonderegger, A., & Sauer, J. (2021). Questionnaire experience and the hybrid system usability scale: Using a novel concept to evaluate a new instrument. *International Journal of Human Computer Studies*, *147*, 102575. https://doi.org/10.1016/j.ijhcs.2020.102575

Brooke, J. (1996). A 'quick and dirty' usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 207–212). Taylor & Francis.

Brooke, J. (2013). SUS: A retrospective. *Journal of Usability Studies*, *8*(2), 29–40.

Cañal, P., Criado, A. M., García-Carmona, A., & Muñoz-Franco, G. (2016). Concepciones didácticas y práctica docente. In P. Cañal, G. Travé González, F. J. Pozuelos Estrada, A. M. Criado, & A. García-Carmona (Eds.), *La enseñanza sobre el medio natural y social. Investigaciones y experiencias2* (pp. 177–205). Díada.

Capps, D. K., Crawford, B. A., & Constas, M. A. (2012). A review of empirical literature on inquiry professional development: Alignment with best practices and a critique of the findings. *Journal of Science Teacher Education*, *23*(3), 291–318. https://doi.org/10.1007/s10972-012-9275-2

Chichekian, T., Shore, B. M., & Tabatabai, D. (2016). First-year teachers' uphill struggle to implement inquiry instruction: Exploring the interplay among self-efficacy, conceptualizations, and classroom observations of inquiry enactment. *SAGE Open*, *6*(2), 215824401664901. https://doi.org/10.1177/2158244016649011

Crawford, B. A. (2014). From inquiry to scientific practices in the science classroom. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of research on science education* (Vol. II, pp. 515–541). Routledge.

de Jong, T., Lazonder, A. W., Chinn, C. A., Fischer, F., Gobert, J., Hmelo-Silver, C. E., Koedinger, K. R., Krajcik, J. S., Kyza, E. A., Linn, M. C., Pedaste, M., Scheiter, K., & Zacharia, Z. C. (2023). Let's talk evidence—The case for combining inquiry-based and direct instruction. *Educational Research Review*, *39*(November 2022), 100536. https://doi.org/10.1016/j.edurev.2023.100536

Fang, S. C. (2020). Towards scientific inquiry in secondary earth science classrooms: Opportunities and realities. *International Journal of Science and Mathematics Education*, *19*, 771–792. https://doi.org/10.1007/s10763-020-10086-6

García-Carmona, A. (2020). From inquiry-based science education to the approach based on scientific practices: A critical analysis and suggestions for science teaching. *Science and Education*, *29*(2), 443–463. https://doi.org/10.1007/s11191-020-00108-8

García-Carmona, A., Criado, A. M., & Cruz-Guzmán, M. (2018). Prospective primary teachers' prior experiences, conceptions, and pedagogical valuations of experimental activities in science education. *International Journal of Science and Mathematics Education*, *16*(2), 237–253. https://doi.org/10.1007/s10763-016-9773-3

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures*, *14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

Hoadley, C., & Campos, F. C. (2022). Design-based research: What it is and why it matters to studying online learning. *Educational Psychologist*, *57*(3), 207–220. https://doi.org/10.1080/00461520.2022.2079128

International Organization for Standarization (ISO). (2018). *Ergonomics of human-system interaction—Part 11: Usability: Definitions and concepts (ISO 9241-11:2018)*.

Ioannidis, J. P. A. (2019). What have we (not) learnt from millions of scientific papers with p values? *American Statistician*, 73(sup1), 20–25. https://doi.org/10.1080/00031305.2018.1447512

Knapp, H. (2018). *Intermediate statistics using SPSS*. Sage Publications, Inc.

Krämer, P., Nessler, S. H., & Schlüter, K. (2015). Teacher students' dilemmas when teaching science through inquiry. *Research in Science and Technological Education*, *33*(3), 325–343. https://doi.org/10.1080/02635143.2015.1047446

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

Lee, S. W. Y., & Tsai, C. C. (2013). Technology-supported learning in secondary and undergraduate biological education: Observations from literature review. *Journal of Science Education and Technology*, *22*(2), 226–233. https://doi.org/10.1007/s10956-012-9388-6

Lewis, J. R. (2014). Usability: Lessons learned… and yet to be learned. *International Journal of Human-Computer Interaction*, *30*(9), 663–684. https://doi.org/10.1080/10447318.2014.930311

Lewis, J. R. (2018). The system usability scale: Past, present, and future. *International Journal of Human-Computer Interaction*, *34*(7), 577–590. https://doi.org/10.1080/10447318.2018.1455307

Liou, P. Y. (2021). Students' attitudes toward science and science achievement: An analysis of the differential effects of science instructional practices. *Journal of Research in Science Teaching*, *58*(3), 310–334. https://doi.org/10.1002/tea.21643

LOMLOE. (2020). Ley Orgánica 3/2020, de 29 de diciembre, por la que se modifica la Ley Orgánica 2/2006, de 3 de mayo, de Educación. Boletín Oficial del Estado, 340, de 30 de diciembre de 2020 [Organic Law 3/2020, of December 29, 2020, which amends Organic Law 2/2006, of May 3, 2006, on Education. Official State Gazette, 340, of December 30, 2020.].

Ly, A., Stefan, A., van Doorn, J., Dablander, F., van den Bergh, D., Sarafoglou, A., Kucharský, S., Derks, K., Gronau, Q. F., Raj, A., Boehm, U., van Kesteren, E. J., Hinne, M., Matzke, D., Marsman, M., & Wagenmakers, E. J. (2020). The Bayesian methodology of Sir Harold Jeffreys as a practical alternative to the P value hypothesis test. *Computational Brain and Behavior*, *3*(2), 153–161. https://doi.org/10.1007/s42113-019-00070-x

Ma, Y. (2023). The effect of inquiry-based practices on scientific literacy: The mediating role of science attitudes. *International Journal of Science and Mathematics Education*, *21*(7), 2045–2066. https://doi.org/10.1007/s10763-022-10336-9

Marrero-Galván, J. J., & Hernández-Padrón, M. (2022). The importance of virtual reality in STEM education: A systematic review from the point of view of experimentation in the classroom. *Bordon*, *74*(4), 45–63. https://doi.org/10.13042/Bordon.2022.94179

MINEDUC. (2012). Bases curriculares. Educación Básica [Bases curriculares. Educación Básica].

NGSS Lead States. (2013). *The next generation science standards: For states, by states* (Vol. 50, pp. 77–82). The National Academies Press. https://doi.org/10.1016/j.endm.2015.07.014

Oliveira, A., Feyzi Behnagh, R., Ni, L., Mohsinah, A. A., Burgess, K. J., & Guo, L. (2019). Emerging technologies as pedagogical tools for teaching and learning science: A literature review. *Human Behavior and Emerging Technologies*, *1*(2), 149–160. https://doi.org/10.1002/hbe2.141

Osborne, J. (2014). Scientific practices and inquiry in the science classroom. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of research on science education* (Vol. II, pp. 579–599). Routledge.

Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, *14*, 47–61. https://doi.org/10.1016/j.edurev.2015.02.003

Perugini, M., Gallucci, M., & Costantini, G. (2018). A practical primer to power analysis for simple experimental designs. *International Review of Social Psychology*, *31*(1), 1–23. https://doi.org/10.5334/IRSP.181

Plomp, T. (2013). Educational design research: An introduction. In T. Plomp & N. Nieveen (Eds.), *Educational design research—Part A: An introduction*. SLO.

Pool, J., & Laubscher, D. (2016). Design-based research: Is this a suitable methodology for short-term projects? *Educational Media International*, *53*(1), 42–52. https://doi.org/10.1080/09523987.2016.1189246

Potkonjak, V., Gardner, M., Callaghan, V., Mattila, P., Guetl, C., Petrović, V. M., & Jovanović, K. (2016). Virtual laboratories for education in science, technology, and engineering: A review. *Computers and Education*, *95*, 309–327. https://doi.org/10.1016/j.compedu.2016.02.002

Raman, R., Achuthan, K., Nair, V. K., & Nedungadi, P. (2022). Virtual laboratories—A historical review and bibliometric analysis of the past three decades. In *Education and information technologies* (Vol. 27, Issue 8, pp. 11055–11087). Springer US. https://doi.org/10.1007/s10639-022-11058-9

Reeves, S. M., & Crippen, K. J. (2021). Virtual laboratories in undergraduate science and engineering courses: A systematic review, 2009–2019. *Journal of Science Education and Technology*, *30*(1), 16–30. https://doi.org/10.1007/s10956-020-09866-0

Reginald, G. (2023). Teaching and learning using virtual labs: Investigating the effects on students' self-regulation. *Cogent Education*, *10*(1), 1–14. https://doi.org/10.1080/2331186X.2023.2172308

Romero-Ariza, M., Quesada, A., Abril, A. M., Sorensen, P., & Oliver, M. C. (2019). Highly recommended and poorly used: English and Spanish science teachers' views of inquiry-based learning (IBL) and its enactment. *EURASIA Journal of Mathematics, Science and Technology Education*, *16*(1), 1–16. https://doi.org/10.29333/ejmste/109658

Rönnebeck, S., Bernholt, S., & Ropohl, M. (2016). Searching for a common ground – A literature review of empirical research on scientific inquiry activities. *Studies in Science Education*, *52*(2), 161–197. https://doi.org/10.1080/03057267.2016.1206351

Rutten, N., Van Joolingen, W. R., & Van Der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers & Education Education*, *58*(1), 136–153. https://doi.org/10.1016/j.compedu.2011.07.017

Scalise, K., Timms, M., Moorjani, A., Clark, L., Holtermann, K., & Irvin, P. S. (2011). Student learning in science simulations: Design features that promote learning gains. *Journal of Research in Science Teaching*, *48*(9), 1050–1078. https://doi.org/10.1002/tea.20437

Strat, T. T. S., Henriksen, E. K., & Jegstad, K. M. (2023). Inquiry-based science education in science teacher education: A systematic review. *Studies in Science Education*, *60*(2), 191–249. https://doi.org/10.1080/03057267.2023.2207148

Sundararajan, N. K., & Adesope, O. (2020). Keep it coherent: A meta-analysis of the seductive details effect. *Educational Psychology Review*, *32*(3), 707–734. https://doi.org/10.1007/s10648-020-09522-4

Toma, R. B. (2022a). Confirmation and structured inquiry teaching: Does it improve students' achievement motivations in school science? *Canadian Journal of Science, Mathematics and Technology Education*, *22*(1), 28–41. https://doi.org/10.1007/s42330-022-00197-3

Toma, R. B. (2022b). Effect of confirmation and structured inquiry on attitudes toward school science. *School Science and Mathematics*, *122*(1), 16–23. https://doi.org/10.1111/ssm.12505

Toma, R. B. (2022c). Elementary school students' interests and attitudes towards biology and physics. *Journal of Biological Education*, 1–12. https://doi.org/10.1080/00219266.2022.2147208

Toma, R. B., Greca, I. M., & Meneses-Villagrá, J. Á. (2017). Dificultades de maestros en formación inicial para diseñar unidades didácticas usando la metodología de indagación [Difficulties of teachers in initial training to design didactic units using the inquiry methodology]. *Revista Eureka Sobre Enseñanza y Divulgación de Las Ciencias*, *14*(2), 442–457. https://doi.org/10.25267/rev_eureka_ensen_divulg_cienc.2017.v14.i2.11

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Marsman, M., Matzke, D., Komarlu Narendra Gupta, A. R., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E. J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin and Review*, *28*, 813–826. https://doi.org/10.3758/s13423-020-01798-5

Vlachogianni, P., & Tselios, N. (2022). Perceived usability evaluation of educational technology using the system usability scale (SUS): A systematic review. *Journal of Research on Technology in Education*, *54*(3), 392–409. https://doi.org/10.1080/15391523.2020.1867938

Vlachopoulos, D., & Makri, A. (2017). The effect of games and simulations on higher education: A systematic literature review. *International Journal of Educational Technology in Higher Education*, *14*(22), 1–33. https://doi.org/10.1186/s41239-017-0062-1

Vorholzer, A., & von Aufschnaiter, C. (2019). Guidance in inquiry-based instruction–an attempt to disentangle a manifold construct. *International Journal of Science Education*, *41*(11), 1562–1577. https://doi.org/10.1080/09500693.2019.1616124

Wieman, C. E., Adams, W. K., & Perkins, K. K. (2008). PhET: Simulations that enhance learning. *Science*, *322*(5902), 682–683. https://doi.org/10.1126/science.1161948

Yánez-Pérez, I., Toma, R. B. Y., & Meneses-Villagrá, J. Á. (2024a). Diseño y usabilidad de IndagApp: una app para la enseñanza de las ciencias por indagación [Design and usability of IndagApp: an app for inquiry-based science education]. *RIED-Revista Iberoamericana de Educación a Distancia*, *27*(2), 223–245. https://doi.org/10.5944/ried.27.2.39109

Yánez-Pérez, I., Toma, R. B. Y., & Meneses-Villagrá, J. Á. (2024b). The IndagApp mobile app—An inquiry-based science teaching resource: Usability evaluation with pre-service teachers. *Journal of New Approaches in Educational Research*, *13*(7), 1–20. https://doi.org/10.1007/s44322-024-00008-7

Yaron, D., Karabinos, M., Lange, D., Greeno, J. G., & Leinhardt, G. (2010). The ChemCollective—Virtual labs for introductory chemistry courses. *Science*, *328*(5978), 584–585. https://doi.org/10.1126/science.1182435

Zacharia, Z. C., Manoli, C., Xenofontos, N., de Jong, T., Pedaste, M., van Riesen, S. A. N., Kamp, E. T., Mäeots, M., Siiman, L., & Tsourlidaki, E. (2015). Identifying potential types of guidance for supporting student inquiry when using virtual and remote labs in science: A literature review. *Educational Technology Research and Development*, *63*(2), 257–302. https://doi.org/10.1007/s11423-015-9370-0

Zeivots, S., Buchanan, J. D., & Pressick-Kilborn, K. (2023). Pre-service teachers becoming researchers: The role of professional learning groups in creating a community of inquiry. *Australian Educational Researcher*, *51*(2), 463–480. https://doi.org/10.1007/s13384-022-00589-2

How to cite this article: Yánez-Pérez, I., Toma, R. B., & Meneses-Villagrá, Jesús Ángel (2024). Design and usability evaluation of a mobile app for elementary school inquiry-based science learning. *School Science and Mathematics*, 1–12. https://doi.org/10.1111/ssm.18303