Computer Science Education

ISSN: 0899-3408 (Print) 1744-5175 (Online) Journal homepage: www.tandfonline.com/journals/ncse20

Computer science camps to reduce the gender gap: a proposal for effective design and impact evaluation

Francesco Faenza, Riccardo Mescoli, Linda Burchiellaro & Claudia Canali

To cite this article: Francesco Faenza, Riccardo Mescoli, Linda Burchiellaro & Claudia Canali (15 May 2026): Computer science camps to reduce the gender gap: a proposal for effective design and impact evaluation, Computer Science Education, DOI: 10.1080/08993408.2026.2671637

To link to this article: https://doi.org/10.1080/08993408.2026.2671637

COMPUTER SCIENCE EDUCATION
https://doi.org/10.1080/08993408.2026.2671637

Routledge
Taylor & Francis Group

REVIEW ARTICLE

OPEN ACCESS Check for updates

Computer science camps to reduce the gender gap: a proposal for effective design and impact evaluation

Francesco Faenza [a], Riccardo Mescoli [b], Linda Burchiellaro [b] and Claudia Canali [a]

[a]Department of Engineering, 'Enzo Ferrari' University of Modena and Reggio Emilia, Modena, Italy; [b]Department of Physics, Informatics and Mathematics University of Modena and Reggio Emilia, Modena, Italy

ABSTRACT

Background and Context: According to recent data, women are still severely underrepresented in Computer Science (CS) in both European and American job markets. To attract female students to CS, several extracurricular initiatives have been started. However, the lack of shared formats and evaluation methods makes it difficult to design effective CS programs and evaluate their impacts.
Objectives: We propose an evaluation tool to assess the impact of the main camp features on the participants, with a particular focus on how these experiences influence female students' future study and career plans. We introduce a guideline for adapting and applying this tool in specific contexts, providing a structured approach to evaluating and refining CS education camps.
Method: We collected relevant literature to synthesize an analysis tool and develop a procedure for evaluating the outcomes of outreach activities. We defined an analysis procedure and provided ready-to-use notebooks for assessing the actual impact of camp activities through regression analysis.
Findings: The resulting tool streamlines the evaluation process, covering every phase from data collection to analysis. It guides the analyst in refining and verifying the results' validity before proceeding with the final analysis. The procedure highlights the most effective camp design choices to trigger positive attitudes among female students towards CS.
Implications: Implementing activities that expose girls to computer science is not sufficient in making them aspire to work or study in this field. Teachers and educators need practical guidance on evaluating their design choices to implement effective education camps and measure their impacts.

ARTICLE HISTORY
Received 2 January 2024
Accepted 6 May 2026

KEYWORDS
computer science education; extracurricular activities; empirical design; gender gap; evaluation tools

1. Introduction

In Europe, the 2022 Digital Economy and Society Index (DESI) (European Commission, 2022) evidences the lack of specialists in technological and engineering fields along with

CONTACT Francesco Faenza francesco.faenza@unimore.it Department of Engineering, 'Enzo Ferrari' University of Modena and Reggio Emilia, Via Pietro Vivarelli 10, Modena, 41125, MO, Italy

Supplemental data for this article can be accessed online at https://doi.org/10.1080/08993408.2026.2671637

a severe gender gap related to these fields: according to the data, only 19% of ICT (Information and Communications Technology) specialists and one in three specialists in the fields of science, technology, engineering and/or mathematics (STEM) graduates are women. The DESI index, which provides a valuable framework for tracking progress toward the EU's objectives in the digital economy and society and underscores the importance of investing in human capital development to ensure long-term growth and competitiveness, warns explicitly about the lack of competencies in the digital fields that the European Union (EU) would experience by 2030. Similar data are shown by the National Science Foundation (National Science Foundation, 2021): only 20% of bachelor graduates in Computer Science (CS) are women. Furthermore, the underrepresentation of women in CS has worsened rather than improved in the last few decades (Sax et al., 2017).

The need to increase the number of CS specialists and to close the related gender gap led schools and institutions to design and implement activities such as education camps and other extracurricular activities as one of the primary methods for encouraging young female students to develop skills and interests in computer science.

Previous research (Heaverlo et al., 2013) has shown that science and math camps and high school learning experiences will link and augment students' interest in scientific fields, suggesting an increase in the availability of similar experiences for the youth. Moreover, the most recent recommendations of the European Commission suggest creating opportunities for schools and educators to attract students to digital careers by promoting participation in initiatives such as CS-based extracurricular activities (European Commission, 2023) or outreach camps, which represent the most common format of informal learning to promote the development of students' interest towards scientific disciplines (Binns et al., 2016). However, despite the plethora of educational initiatives in this field and available resources to be used during these activities, designing and implementing successful and effective CS programs remains a challenging task, as it requires experience and careful consideration of multiple factors.

Even enlarging the discussion to the STEM (Science, Technology, Engineering, Math) disciplines and despite all the attention that education activities in that context recently received (Li et al., 2020; Martn-Páez et al., 2019), having a common instrument of evaluation has been highlighted in the literature as of paramount importance (Decker et al., 2016; Rorrer, 2016). A further important point is the presence of evaluation tools specifically designed for the unique characteristics and goals of a STEM or CS education camp for female students (Fronza et al., 2020; Zailan et al., 2019). The absence of appropriate evaluation tools makes it challenging to measure the actual impacts and effectiveness of the implemented activities and, consequently, to refine and improve the design of future initiatives. Some examples of previous research attempts have focused on providing a library of ready-to-use surveys (Decker & McGill, 2019) from various contexts with the objective of having a library extensive enough to cover as many specific cases as possible or creating an "a la carte" survey (Rorrer, 2016) to give researchers enough questions to build the particular survey of their needs.

The contribution of this paper is twofold. First, we develop an evaluation tool based on our own experience and on the existing literature to assess the effectiveness and impact of CS education camps. This tool enables future educators to systematically evaluate key camp features and their influence on participants. However, as highlighted by Knekta et al. (2019) in their influential work, using a questionnaire "as is" can be problematic:

validation in the particular context of the study is fundamental. Therefore, our tool is not intended as a static instrument, but as a flexible instrument that supports context-aware adaptation. With this evaluation tool in place, we then turn our attention to introducing a guideline for its adaptation and application in specific contexts, providing a structured approach to guiding the design and assessment of CS outreach camps. Specifically, the evaluation tool was developed starting in 2016 within the context of *Digital Girls* educsci11110715, FaenzaRIIForum (Faenza et al., 2021a, 2021b, Faenza et al., 2021c), a long-running Italian project inaugurated in 2013, aimed at engaging female students aged 16–18 and attracting them to computer science. The initiative has been included in the Case Study Library of the Observatory for Public Sector Innovation (OPSI, 2023) and mentioned in the 2021 She Figures report.[1] The resulting survey was then tested in the context of the "STEM for Future" ERASMUS+ project[2] and the 2022 edition of Digital Girls.

We demonstrate how to effectively use, refine, validate, and analyze an evaluation survey using our tool to identify the strengths and weaknesses of different teaching methods and activities, other than verifying intended outcomes. Our goal is to equip educators with practical guidance for analyzing data collected through the proposed evaluation tool. We provide a comprehensive tool that includes a base survey ready to be imported, used, refined, and analyzed alongside included Jupyter notebooks. These notebooks facilitate the validation and reliability assessment of the instrument, data import and cleaning, and data analysis. Building upon this initial contribution, the proposed tool has evolved into a deployable platform that facilitates its practical adoption by outreach organizers and researchers (Burchiellaro et al., 2025). We offer a detailed description of how to use the tool, taking a significant step toward the standardization called for in the literature. This tool is designed to support researchers from the survey's definition to the analysis of its results, enhancing the overall rigor and effectiveness of CS camp evaluations. All materials necessary for reproducing the suggested assessment and analysis are available at the public GitLab repository.[3]

The rest of this paper is structured as follows. Section 2 considers some related work about existing classifications of STEM camps and related assessment tools, while in Section 3 we present the proposed evaluation tool. In Section 4, we identify the main features that can be used for CS camps analysis. Section 5 presents a use case of the tool that guides the entire analysis process after the data collection phase, while Section 6 discusses the implications of our findings for future research and practice in CS education. Finally, we conclude the paper with some final remarks.

2. Background

2.1. Characterization of outreach camps

To the best of our knowledge, while there have been efforts to classify extracurricular activities in computer science education, existing approaches often lack a standardized and easily applicable framework for educators. Much of the literature on CS education primarily focuses on curricular aspects, such as foundational concepts and programming languages, rather than on the unique role of extracurricular activities in attracting students to the discipline. Prior studies Decker and McGill (2019); McGill and Decker (2020) have highlighted gaps in the systematic evaluation and classification of computing education initiatives, particularly in pre-college contexts. Specifically, the literature does not provide a consolidated analysis of the structural and organizational characteristics of computer science outreach camps that could be directly used to support systematic evaluation. As the identification of such characteristics is a prerequisite for analyzing and comparing outreach activities, and given the limited availability of CS-specific studies focusing explicitly on these aspects, we extend our scope to STEM-based outreach programs. The broader STEM education literature offers several studies that explicitly examine the key features of extracurricular and outreach initiatives, which are largely applicable to computer science contexts.

In this section, we aim to identify the main features driving the design choices for the implementation of education outreach camps with a specific focus on counteracting the gender gap in computer science. The study in Guzey et al. (2016) presents the STEM Integrated Curriculum Assessment Tool (STEM-ICA), specifically designed to measure the effectiveness of STEM program design. Based on prior studies, the STEM-ICA comprises nine main features, including motivating and engaging context, engineering design, integration of science and mathematics, instructional strategy, teamwork, communication, assessment, and organization. The research in Zheng et al. (2022) develops an indicator system to assess STEM teaching cases. The authors evaluate 51 STEM teaching cases conducted across China, incorporating indicators derived from international literature. In Lynch et al. (2018), authors focus on Inclusive STEM High Schools (ISHSs) in America, deriving ten main features characterizing exemplar cases of ISHSs through literature analysis and what authors term “excellence case evaluation”. In Nite et al. (2017), a comprehensive classification of STEM activities is offered by identifying and synthesizing common STEM teaching and learning characteristics through an extensive literature review. The findings are aggregated into four main categories: student-centered practices, informal venues, teacher factors, and technology adoption, with additional key indicators within each category. A clearly outlined framework for instructional strategies in integrated STEM within secondary education is presented in Thibaut et al. (2018). This framework encompasses five categories of instructional elements, consistently identified in the selected literature as crucial for effectively teaching integrated STEM. Finally, the authors in Martn-Páez et al. (2019) investigate the implementation of educational interventions self-identified as STEM, revealing discrepancies in the theoretical framework and advocating for the adoption of a vision that seamlessly integrates STEM disciplines.

Collectively, these works provide valuable insights into the key characteristics of extracurricular STEM and CS activities, which, as we will illustrate in the following sections, serve as a foundation for developing a robust evaluation system. Understanding these characteristics is crucial as they define aspects of the initiatives that should be assessed to determine their effectiveness. By evaluating these features, institutions conducting multiple iterations of the same program or operating across different venues can assess whether specific design choices have been successful or whether alternative approaches may yield better outcomes for targeted objectives. In Section 4, we will outline how these characteristics inform the construction of an evaluation tool, integrating findings from existing studies with our direct experience in the field.

2.2. Evaluation tool

Evaluating an extracurricular education camp aimed at reducing the gender gap in computer science significantly differs from assessing a standard curricular course, as the primary focus is usually more on factors such as camp satisfaction, engagement, and orientation rather than just on technical skills and achievements. In promoting CS, various initiatives have been implemented over the past decades. However, a common challenge persists: the lack of a proper evaluation tool to assess the actual impact and benefits of these initiatives.

A recent study Faenza, Canali, Colajanni, et al. (2021c) analyses several CS-related initiatives and points out a noticeable scarcity of published data on their assessed outcomes: this lack hinders a comprehensive understanding of the effectiveness and success of such programs. On the other hand, a structured, specifically designed assessment tool may be helpful to enable educators and organizations to assess the effectiveness of their programs. Such assessment is particularly important in informal and extracurricular contexts, where outcomes are often indirect, long-term, and strongly mediated by contextual (Craig, 2016; Knekta et al., 2019) A precise assessment can foster an evidence-based design practice, while the availability of design choices and lesson logs could serve as a foundation for organizations to develop or reproduce a CS program. Many structured assessment tools exist; however, they often lack direct applicability to outreach settings or require substantial adaptation to align with the specific goals, constraints, and audiences of CS camps.

This gap has been widely recognized in computing education research. For example, Wilson et al. (2013) examined the role of institutions in defining extracurricular activities for STEM students, discussing how institutional contexts influence student participation and the perceived value of these activities. Their findings highlight the importance of structured evaluation in understanding how different institutional settings impact outreach efforts. Similarly, McGill and Decker (2020) conducted a gap analysis to improve research and experience reports of pre-college computing activities, emphasizing the need for standardized reporting practices to enhance comparability and effectiveness across different programs. Additionally, Decker et al. (2016) advocates for a common framework for evaluating computing outreach activities, aiming to establish consistent assessment methods and improve the quality of outreach efforts. These studies underline the necessity of an evaluation framework that accounts for the unique characteristics of extracurricular CS initiatives. Rorrer (2016) reached a similar conclusion, proposing the CISE REU evaluation toolkit – an "a la carte" solution for researchers looking to develop tailored evaluation tools. Decker and McGill continued this work with an insightful literature review of evaluation tools (Decker & McGill, 2019), which served as the foundation for a library of evaluation instruments hosted on csedresearch.org. This database provides a filterable list of surveys, enabling researchers to find the most appropriate tools for their needs. Among the reported characteristics for each tool is an indication of whether there is evidence of reliability and/or validity.

A notable contribution is the work by Craig (2016), which proposes a theory-driven evaluation framework for gender and computing interventions. Craig (2016) explicitly argues that intervention outcomes cannot be meaningfully assessed without articulating the underlying assumptions and contextual conditions under which change is expected to occur. At the same time, the framework intentionally avoids prescribing fixed instruments, acknowledging that evaluation tools must be instantiated and adapted to local contexts. This position is consistent with the widely recognized limitation that "one size does not fit all" in educational evaluation, particularly in informal and outreach settings, as advocated by the work of Knekta et al. (2019).

Other studies further illustrate this challenge. For example, in Danoff (2017), the authors present a methodology for assessing gender barriers in Computer Science at Harvard, but the focus is on college students in the specific context of the Harvard faculty. Another example is the work in Burge et al. (2013), which details the outcomes of the "Girls on the Go" program. The authors' main focus is to present results and recommendations, and pieces of information about the survey structure are discernible only indirectly through their analysis section. In contrast, in Davis and Hardin (2013), authors contribute by providing a specific and detailed account of their experiences as organizers of STEM camps, highlighting the best practices of camp implementation. Additionally, studies such as Yilmaz et al. (2009) and Mohr-Schroeder et al. (2014) offer the point of view of educational camp organizers, furnishing instructional plans, class particulars, and survey proposals. However, it is noteworthy that the assessment tools in these works are presented in a somewhat peripheral way, often in the form of generic suggestions.

In the case of Spieler et al. (2020), the authors conducted an extensive literature review of programs aimed at reducing the gender gap in CS, highlighting the main factors influencing the intention to pursue a degree course in CS, thus indirectly providing insights for defining a proper assessment tool for such programs. Interestingly, the study conducted in Vrieler et al. (2021) emerges as the closest match to our objective, delving into gender differences among members of a computer science club. Notably, the authors include the proposed questions in the appendix and meticulously elucidate their analyses. Although the primary focus of the study is not to provide a ready-to-use questionnaire for camp organizers, it nevertheless offers valuable resources to support our research objectives. It is worth noting that the study drew on a broader work by Archer et al. (2015), where the authors attempted to define a methodology for calculating a "science capital" score.

Taken together, these studies highlight a clear methodological gap: while frameworks and recommendations for evaluating CS outreach exist, there is limited support for translating them into deployable, adaptable, and validated assessment instruments that respect contextual variability. Our assessment tool, which will be presented in Section 3 along with suggestions for its application, has been refined over time, incorporating valuable insights from the literature, such as the significance of participants' self-perception about capabilities in CS (Decker & McGill, 2019; Lewis et al., 2011; Rorrer, 2016).

3. Assessment tool for CS camps

In this section, we introduce the assessment tool that we developed and refined over the last few years of research on this topic. A preliminary version of the tool, along with a description of its development process, was presented in Faenza and Canali (2023). This description highlights the evaluation tool's structure and details, emphasizing that it represents a flexible base template to be adapted according to the specific context and not a one-size-fits-all solution. It is worth noting that our process of tool refinement is akin to design research (Cobb et al., 2015), a methodology outlined in the literature that allows researchers to systematically reflect on their experiences and improve their work through a structured research process. In fact, the work embraced some of the principles of design-based research (Reimann, 2010), incorporating an iterative design process informed by insights from a literature review, including and adapting new tools for testing purposes and validating the result with focus group sessions, enabling a comprehensive exploration of critical aspects.

As discussed in Section 2, our design-based research approach drew inspiration from both generalized STEM instruments and CS-specific tools. The resulting survey was influenced by the Science Capital questionnaire developed by Archer et al. (2015) and the survey study conducted by Vrieler et al. (2021), where the Archer questionnaire was specifically adapted to the context of CS clubs. Additionally, the work of Decker and McGill (2019) was helpful in reviewing existing evaluation tools in computer science. While their work focuses on cataloging and classifying available instruments, our approach differs by aiming to provide a flexible, general-purpose tool that can be readily adapted and used to design, administer, and analyze an extracurricular activity's impact. This aligns with Decker and McGill's emphasis on the need for a common adaptable evaluation framework Decker et al. (2016).

Given the substantial differences in evaluation between curricular and extracurricular initiatives, we focused on program evaluation tools that assess non-cognitive aspects, as opposed to the primary emphasis on conceptual understanding and knowledge acquisition typically found in curricular settings. In extracurricular activities, while content knowledge may still be relevant, factors such as engagement, motivation, self-efficacy, perceptions of the field and future academic or career intentions play a more central role in evaluation. Among the many examples provided in the review of Decker and McGill (2019) that proved useful, the FOSS2Serve project[4] was of particular interest, focusing on assessing the usefulness and impact of solving real-world problems. Braswell et al. (2021) shared common indicators, other than the emphasis on girls, while Rorrer (2016) provided a detailed explanation of the procedure that led to their proposal of a generalized solution for CS evaluation. These resources collectively informed the development of our survey proposal for the evaluation tool.

Our research group's substantial experience primarily centers on developing and implementing CS education activities for secondary school students. Consequently, the survey's structure and content are tailored to align with the characteristics and needs of this specific target audience. Hence, while the work by Archer et al. (2015) is used as an important source for items, its primary focus on defining a science capital score does not align with our specific objectives. Furthermore, it is important to note that our study neither emulates the approach of Vrieler et al. (2021). In contrast to their aim of assessing the existing background, our study seeks to evaluate the impact of activities. Nonetheless, the work by Vrieler et al. (2021) remains a valuable source of insight throughout our development process, as we share a common focus on K-12 students and CS activities. Moreover, our study does not repeat any of the aforementioned existing CS questionnaires. While we share a common interest in evaluating impact, particularly on minorities, our study proposes an entire analysis tool rather than a starting set of questions. Still, investigating the questions and their significance, with appropriate references to the literature, will help researchers optimize their use of the tool and guide them in adapting the proposed set of questions to their specific needs. Indeed, we emphasize that while the primary focus of our research is CS, the proposed evaluation tool can be readily adapted to the broader STEM domain, as will be explained in the subsequent sections.

Our proposal of an evaluation tool is organized into six distinct sections: Design and structure, Subscription Data, Pre-Survey questions, Post-Survey questions, Reliability and Validity testing, Data analysis preparation, and Technical aspects. A concise summary of the survey's key points is presented in Table 2. Additionally, the complete list of detailed questions and an importable version of the survey designed for LimeSurvey[5] have been made available at the public GitLab link[6] as well as the related notebooks for reliability and analysis. Building on this initial set of instruments, the evaluation tool has evolved into a deployable platform that integrates survey generation, data validation, and analysis workflows, with the goal of simplifying adoption by educators and outreach organizers (Burchiellaro et al., 2025). In line with privacy and data protection requirements, the platform is designed to be self-hosted by the deploying institution, while an online instance is provided exclusively for demonstrative and exploratory purposes. The aim is to ensure transparency, reproducibility, and practical usability, while fostering a collaborative community around the tool. As our research progresses, we continuously refine and expand the tool, and we hope that by making it publicly accessible, researchers and practitioners will contribute to its evolution. Our vision is to develop a tool that can be easily adapted to different contexts and remains user-friendly for individuals with varying levels of expertise. For clarity, we will use "impact evaluation procedure" to refer to the overall methodological workflow guiding the design, adaptation, validation, and analysis of outreach camp evaluations. The term "evaluation tool" will refer to the complete set of supporting artifacts provided, including the survey instrument, analysis notebooks, and technical infrastructure.

3.1. Design and structure

The development of the proposed survey was guided, as stated, by a research methodology akin to design research (Cobb et al., 2015). The full process comprised successive iterations of literature review and integration, successive refinement through pilot testing and other techniques for in-depth research of stimuli to propose, such as focus groups. The literature review process served as a foundational step, allowing us to identify and validate key factors that influence participants' attitudes, perceptions, and future intentions related to STEM and CS fields. This review drew from a wide array of studies on education, gender disparities in STEM and psychological factors, ensuring that the questionnaire was both theoretically grounded and practically relevant. Moreover, as a further check, synthesized questions were compared and eventually modified according to other research results in the direction of refining a survey for CS outreach initiatives. As stated in the introduction of this section, many other attempts were consulted and compared to our proposal, confirming the correspondence both in the general methodology applied, further described ahead in this section, and the main indicators behind what is being asked, described in the Pre- and Post-Survey subsections.

Given the transformative nature of initiatives like these, to further confirm the main indicators behind the questions being asked, we also employed focus groups with past participants in our program and female CS undergraduates. These focus groups provided valuable qualitative insights into the participants' experiences, expectations, and challenges. Feedback from these sessions led to adjustments in the survey structure, such as the inclusion of questions related to parental influence and video game engagement, which were identified as relevant factors for understanding students' prior exposure to technology.

To further ensure the reliability and validity of the survey, pilot tests were conducted at each iteration with a smaller subset of the target population. These tests were instrumental in identifying any ambiguities or biases in the questions, leading to additional refinements.

The proposed research design encompasses a comprehensive assessment tool structured into distinct Pre-Survey and Post-Survey phases. The Pre-Survey phase is meticulously designed to delve into the participants' background characteristics. This preliminary investigation serves as a crucial foundation for understanding the baseline conditions of the participants before their engagement with the program. On the other hand, the Post-Survey phase is intricately designed to capture the effects and outcomes derived from the program. Furthermore, recognizing that the preliminary phase could yield an extensive survey, integrating specific background information inquiries into a separate subscription process could be a solution. This design also allows us to investigate any pre-existing contextual characteristics that might have contributed to the participants' initial responses. Another factor supporting our proposal of collecting pertinent background data as part of a separate subscription process is based on the notion that, since participation is usually voluntary, participants may be more inclined to respond to questions with careful consideration. Table 2 summarizes the proposed structure.

3.1.1. Subscription phase

The subscription survey aims to capture generic background information from participants during the sign-up phase, providing essential context for the analysis. The collection of background information encompasses various aspects and greatly depends on the specific context of the extracurricular CS program. This may include essential details such as nationality, school type, age, sex, and town of residence. For instance, in a region with students hailing from diverse areas of the country, data concerning their exact residence place could be particularly important in understanding regional differences and potential influences on the program's outcomes. Similarly, in regions with a solid integration of immigrant populations, data regarding the duration of their stay could provide valuable insights into the potential impact of cultural background on the participants' experiences and engagement. Among the diverse range of background data, some information is universally valuable and can significantly contribute to the analysis. For instance, in the Italian context, in the case of secondary school participants, it is essential to collect data on the type of school to allow differentiation in the analysis between students from high schools, professional institutes, and technical institutes.

3.1.2. Pre-survey phase

Additional personal background information will be collected during the Pre-Survey part of the evaluation tool, starting with more specific questions focusing on the parents' background, considering literature about parental figure influence on the field of study choice (Gabay-Egozi et al., 2015) and the "Social support" section in the work of Vrieler et al. (2021). Depending on the scope of the camp, questions will be posed to understand whether the parents work in the STEM or CS field, their level of passion for STEM or CS themes, and their educational background. Among these questions, particular importance is given to determining the parents' employment field, which holds priority in the event that the survey needs to be shortened. This choice is due to existing literature (Stanko & Zhirosh, 2017) that emphasizes the influence of parents' employment on their children's future intentions. In particular, the study by Stanko and Zhirosh (2017) suggests that parents' careers significantly impact the career plans of young women, particularly in the context of IT. By distinguishing between children with parents in STEM or CS fields and those without, the analysis phase can assess potential differences in the impact of the program.

In the specific context of CS, two other crucial questions are related to participants' engagement with video games and their coding experience (Jenson et al., 2007). Participants will be asked whether they play video games or not, and for those who respond affirmatively, an additional question will be about the number of hours per week they spend playing.

Similarly, participants will be asked about their coding experience, if they have any prior experience, and, if so, to self-evaluate their level of coding proficiency. Related to this question, the literature suggests (Lewis et al., 2016) that participants should be asked to state the self-perception of their CS capabilities and their identification with the perceived identity of a CS expert.

The Pre-Survey section, which focused on future intentions, is of utmost importance Ajzen (2001). The level of granularity in this section can be tailored to suit the length of the intended survey and the research objectives. In the work of Vrieler et al. (2021), being the main focus of the assessment of existing background, this section is extended to investigate in depth future intentions, parental attitude about the plan, and work aspirations. In its shorter version, the pretest can simply include a straightforward question asking participants if they intend to pursue university studies or find a job in the STEM or CS field.

In line with the research focus of our group, the survey includes questions regarding gender stereotypes. We are interested in exploring participants' perspectives on gender-related factors that might impact their engagement and pursuit of CS and STEM disciplines. Finally, in the Pre-Survey, an open-ended question asking what the participant expects to find during the activities could be helpful. This will allow us to better understand the eventual results that are not aligned with the expectations. In addition, we include an open-ended question asking participants to describe, in their own words, what they believe computer science to be. This qualitative baseline allows us to capture participants' initial perceptions of the field, which are often shaped by stereotypes or limited prior exposure (Semmens et al., 2015).

3.1.3. Post-survey phase

In the Post-Survey part of the evaluation tool, our focus remains on assessing the program outcomes and their impact on the participants. To achieve a comprehensive analysis, we believe it is crucial to repeat certain key questions from the Pre-Survey. Specifically, we will revisit the inquiries regarding participants' future intentions in terms of pursuing STEM or CS studies, as well as their perceptions of gender stereotypes in these fields, other than self-perception of their CS/STEM capabilities Bandura and Wessels (1997). By comparing the Pre-Survey and Post-Survey responses, we can gain valuable insights into any changes or shifts in participants' intentions and attitudes that may have occurred as a result of their engagement in the program.

In the final phase of the Post-Survey evaluation, we turn our attention to the assessment of camp outcomes. To measure the effectiveness of the program in enhancing participants' understanding of CS or STEM themes, we include questions that gauge the level of improvement perceived in knowledge and skills related to the field. Our study primarily focuses on students' perceptions, which is a common approach in extracurricular activities. In such contexts, the main goal often does not involve assessing students' progress in acquiring specific, measurable knowledge.

Additionally, the Post-Survey includes assessment items that capture participants' overall experience satisfaction and engagement with the camp. These questions will address aspects such as participants' enjoyment and team belonging. It is essential to highlight the importance of assessing self-efficacy. Defined as an individual's belief in his or her capacity to execute behaviors necessary to produce specific performance attainments (Bandura, 1977), in the field of CS, numerous studies have illustrated the significant correlation between self-efficacy and career orientation (Aivaloglou & Hermans, 2019; Kallia & Sentance, 2018; Rosson et al., 2011).

By analyzing participants' responses, our goal is to comprehensively understand the camp's success and identify areas for potential improvement, thereby refining future extracurricular CS or STEM initiatives. We will illustrate this further with analysis examples in Section 5.

3.2. Reliability and validity testing

In line with established recommendations in educational measurement, which emphasize that validity is context-dependent and must be empirically supported rather than assumed (Craig, 2016; Knekta et al., 2019), the survey instrument was rigorously validated and tested for reliability using a combination of Exploratory Factor Analysis (EFA) and reliability metrics such as Cronbach's alpha.

Exploratory Factor Analysis was first employed to uncover the underlying factor structure of the survey items, identifying key latent constructs that the survey aims to measure. Factors were extracted based on eigenvalues greater than one and confirmed through scree plot inspection and parallel analysis. A Promax rotation method, an oblique rotation that allows factors to be correlated, was applied to clarify the factor structure. Items with significant pattern coefficients (typically above 0.4) were retained, while those with substantial cross-pattern coefficients were reviewed for potential revision or removal.

Following EFA, Principal Component Analysis, or other reduction methods, was conducted to further validate the factor structure by reducing data dimensionality and confirming that the identified factors accounted for a significant portion of the variance.

Cronbach's alpha was calculated for each subset of factors measuring distinct patterns to assess the survey's internal consistency. A Cronbach's alpha value above 0.7 across all factors indicated strong internal reliability, confirming that the items within each factor consistently measured the intended pattern.

Following this iterative procedure throughout the course of our research, we were able to refine specific factors and, as a result, develop a series of notebooks capable of identifying and managing outliers. This careful management of outliers prevented them from influencing the results. Moreover, the factors identified through the process correspond with the question groups presented in Table 2.

This approach directly reflects the methodological guidance proposed by Knekta et al. (2019), which stresses the role of factor analysis and reliability testing in establishing validity evidence for survey instruments used in new or evolving contexts. Consistent with Craig (2016)'s framework for evaluating CS gender-related interventions, the validation process was treated as an integral part of the evaluation procedure rather than as a one-time, fixed property of the instrument. This validation and reliability testing process ensured that the final survey instrument has evidence of reliability and validity in the specific context considered, providing a solid foundation for the subsequent analysis phase, where, as shown in Section 5, regression analysis was used to identify variables influencing key outcomes of interest.

Given that changes in context or modifications to the survey, such as adding or altering questions, may influence the effectiveness of the indicators, it is highly recommended to run an Exploratory Factor Analysis and Cronbach's alpha on the modified or added factors using the provided notebooks. In particular, the "outliers" and "validation" notebooks are designed to assist in this process. An example of how to use these notebooks is provided in the case study section. If the survey is delivered in a similar context and the factors remain consistent, rerunning these analyses may not be mandatory, but it is still recommended. This is particularly important when the survey is translated into the language of the new context, as translation can affect EFA results. It is also important to consider sample size in such cases, as highlighted by Knekta et al. (2019), who found that sample size plays a crucial role in the accuracy of reliability measures.

3.3. Data analysis preparation

For a quasi-experimental repeated measures design, it is crucial to perform specific data treatments before proceeding with analysis to ensure the integrity and accuracy of the results. One essential step is to match Pre-Survey responses with corresponding Post-Survey responses, ensuring that only participants who completed both surveys are included in the analysis. This is vital for maintaining the consistency and reliability of the data.

If you are using the suggested platform, as detailed in the subsection 3.4, it will be straightforward to download only complete responses – those where participants have answered every question from the beginning to the end of the survey. In cases where a different platform is used, it is important to manually filter out incomplete results from both the Pre-Survey and Post-Survey to avoid skewing the analysis.

Another critical consideration is the connection between Pre-Survey and Post-Survey responses. Depending on the platform being used, it is essential to establish a reliable method to link a participant's Pre-Survey responses with their Post-Survey responses. One approach could be asking participants to create and remember a unique code generated by combining part of their name with a random number. However, a more reliable method would be to use a platform like the one we recommend, where it is possible to associate each participant automatically with a unique code generated by the platform. This code would then be used to link the Pre- and Post-Survey data accurately.

If you choose to use the proposed evaluation procedure without adopting the full platform described in burchiellaro2025elevateai, the analysis notebooks can still be used independently. All notebooks are openly available in the public GitLab repository and can be downloaded, adapted, and executed on locally prepared datasets, provided that the input data follow the expected structure. This design choice allows researchers and practitioners to benefit from the proposed analysis workflow even when institutional, technical, or privacy constraints prevent the deployment of the full platform.

All the provided notebooks for analysis, as further explained in Section 5, use CSV file formats as input with specific column names. In the illustrative case, an association between column names and survey questions is provided, making it possible to adapt the notebooks to different instruments beyond the one proposed here. As both the survey instrument and the supporting notebooks are continuously refined, users are encouraged to refer to the GitLab repository for the most up-to-date versions of the materials and documentation.

3.4. Technical aspects

From a technical standpoint, a requirement of the tool is to establish a clear connection between the responses provided by each participant before and after their participation in the program. By doing so, we can assess any potential shifts in attitudes, perspectives, or knowledge as a result of program attendance. Additionally, as stated, generic background data may be collected during the subscription phase. In this case, a key indicator, such as the email address, can be used to connect background data to the Pre-Survey and Post-Survey answers.

Obviously, before initiating any substantial analysis, a critical step in our methodology is the anonymization of the data to ensure participant confidentiality and data protection. In fact, it is important to remember that in the European context, due to data privacy regulations, it is crucial to emphasize the implementation of a clear and comprehensive privacy policy. The participant has to explicitly accept the policy, agreeing to let their data be used in an anonymized and aggregated form.

The choice of survey administration platform is of significant importance in the evaluation process. While widely used platforms such as Google Forms are of common use, our research group encountered specific challenges that necessitated an alternative approach. For our proposed Pre-Survey and Post-Survey design, it was essential to maintain a connection between the responses submitted by participants. To fulfill this requirement, we successfully implemented the design using LimeSurvey, an Open Source platform offering enhanced data control when self-hosted. Notably, LimeSurvey's built-in participant handling mechanism allowed us to assign a unique random code (token) to each participant. Depending on the situation, this token could either remain transparent to the participant through email invitations or be explicitly provided when an email-based solution is not feasible. This coding system facilitated the seamless linking of Pre-Survey and Post-Survey surveys in our analysis instrument. These technical requirements and design choices later motivated the development of a dedicated, self-hostable evaluation platform that generalizes and automates this workflow, as described in burchiellaro2025elevateai.

By leveraging this approach, we were able to ensure the integrity and continuity of the data collected, enabling a more comprehensive assessment of the program's outcomes.

4. Features for CS camps analysis

This section addresses the identification and classification of key features characterizing extracurricular computer science camps. As discussed in Section 2, the computer science education literature lacks studies explicitly dedicated to systematically categorizing outreach camp characteristics. For this reason, we draw on selected studies from the broader STEM education literature in which feature identification and classification constitute an explicit research objective.

In this section, we present a synthesis of the main features identified in the analyzed works on CS educational camps. This synthesis is derived from existing literature and further shaped by insights gained through the design, implementation and assessment of CS camps focused on reducing the gender gap.

The selection of the analyzed studies (Table 1) was guided by their explicit focus on identifying, organizing or systematizing the defining features of extracurricular initiatives. While additional studies report individual characteristics or design choices, these features often emerge only implicitly in outcome analyses and are not the primary object of investigation. In contrast, the selected works provide structured perspectives on program characteristics, making them particularly suitable for comparative analysis.

To synthesize the findings from the analyzed works, we identify five categories of features for extracurricular initiatives, summarized in Table 3: main program characteristics, objectives, knowledge content, activity design and supporting tools. To enhance the readability of the table, the "Research Reference" column lists the IDs defined in Table 1 rather than full citations.

Table 1. Studies about STEM/CS camps classification.

Reference	ID1	Title	Subject
Guzey et al. (2016)	1	Building Up STEM: An Analysis of Teacher-Developed Engineering Design-Based STEM Integration Curricular Materials	Proposal of a STEM Integration Curriculum Assessment (STEM-ICA) to assess the design effectiveness of STEM curriculum programs.
Zheng et al. (2022)	2	K–12 Science, Technology, Engineering, and Math characteristics and recommendations based on analyses of teaching cases in China	Identification of common characteristics of Chinese STEM courses and proposal of a 13-indicator system for evaluation.
Lynch et al. (2018)	3	Understanding inclusive STEM high schools as opportunity structures for underrepresented students: Critical components	Analysis of Inclusive STEM High Schools (ISHSs) in the US, with a focus on critical features of exemplary cases.
Nite et al. (2017)	4	Explicating the Characteristics of STEM Teaching and Learning: A Metasynthesis	Classification of STEM educational activities based on literature review.
Thibaut et al. (2018)	5	Integrated STEM Education: A Systematic Review of Instructional Practices in Secondary Education	Framework for instructional practices in integrated STEM based on systematic literature review.
Martn-Páez et al. (2019)	6	What are we talking about when we talk about STEM education? A review of literature	Literature review on the implementation and theoretical implications of STEM educational activities.

[a]This column defines a numeric ID hereafter used for brevity to reference the work in other tables.

Table 2. Assessment tool structure.

Administering at	Question groups	Questions focus
Subscription survey	Background information	Age, Sex, City, Nationality, School type, Town of living
Pre-Survey	Parents background	Job field (CS/STEM), CS/STEM passion, instruction level
	Participant background	Video games engagement, previous coding experience
	Future intentions	Future intentions regarding study or work
	Gender bias	Perception about gender stereotypes in the field
	Expectations	Participant expectations for the program
Post-Survey	Future intentions	Future intentions regarding study or work
	Gender bias	Perception about gender stereotypes in the field
	Perception shift	Perception shift towards CS disciplines and themes
	Camp evaluation	Overall experience satisfaction and engagement with the camp

Table 3. Features for CS camps analysis.

Category	Feature Name	Research Reference (ID)1
Main program characteristics	Class size	2, 4
	Age range	2, 4
	Race and Gender	6
	Class hours	2, 4, 6
	Program type	4
	Main topics and related subjects	2
	Motivating and engaging context	1
	Inclusiveness	3
	Positive community	3, 6
	Support network	3
	Agency and choice	3
Objectives	Teamwork and communication	1, 5
	Informal learning	3, 6
	Scientific spirit and innovative ability	2
Knowledge content	STEM/CS-focused	2, 3
	Multidisciplinarity	2, 5, 6
	Connection to real-world problems	2, 3, 5, 6
	Student-centered approach	2, 5, 6
Activity Design	Informal venue	3, 4, 6
	Interdisciplinary approach	2, 5, 6
	Engineering design	1, 2, 3, 4, 5
	Teaching methodology	1, 3, 4, 5, 6
	Presentation and reflection	1, 2
	STEAM	2
Tools	Software and hardware	2, 4, 6
	Assessment tools	1, 2, 3, 5
	Integrated technology use	2, 3, 4, 5
	Venue support	2
	Educational resources	1, 2
	Teacher-related factors	2, 3, 4

[a]ID refers to the numeric ID defined in Table 1.

This feature-based synthesis follows the rationale articulated by Craig (2016), which emphasizes the importance of making program assumptions, design choices, and contextual conditions explicit when evaluating educational interventions. By structuring features into coherent categories, we provide a transparent foundation for subsequent analysis and evaluation.

It is worth noting that the studies identified in Section 2 shared significant commonalities, with two aspects consistently emphasized: the importance of interdisciplinarity and the documentation of implemented activities.

Acknowledging computer science's intrinsic multidisciplinary nature (Yasar & Landau, 2003), it's crucial to highlight the valuable role of interdisciplinarity in extracurricular initiatives (Burge et al., 2013; Franklin et al., 2013). In fact, interdisciplinary approaches, which go beyond mere juxtaposition of disciplines (Klein, 2010; Nicolescu, 2014), involve a deeper integration where multiple disciplines are seamlessly combined, benefiting STEM initiatives and significantly enhancing computer science program development.

Finally, most of the works stress the importance of a student-centered approach as related to higher student participation in the lesson process, suggesting approaches like Project-Based Learning. A shared emphasis on evaluating and assessing teaching and learning activities was also observed.

Although some features reported in Table 3 may seem redundant, they emphasize different aspects, as evidenced in the literature. Moreover, a detailed explanation of each feature will be provided in the following subsections. For example, Lynch et al. (2018) identifies both the “inclusive STEM mission” and “agency and choice” features, where the former pertains to the STEM mission that a program should adopt, while the latter relates to the level of support the program should provide to students.

It is worth noting that certain features, although initially modeled by a specific study, are acknowledged and confirmed by other authors. For instance, in Lynch et al. (2018) authors emphasize the aspect of inclusion, translated into relevant features such as “Inclusiveness”. Although not explicitly modeled as a distinct feature by other authors, the same theme of inclusion is addressed in their work. Some complex indicators are intentionally maintained: for instance, Guzey et al. (2016)’s evaluation of teamwork and communication could have been included under a more general soft-skill category. However, given that “collaboration” is a central focus in other studies, as in the study by Zheng et al. (2022), we retain this indicator to emphasize the significance of fostering teamwork and cooperation. On the other hand, the feature “scientific spirit and innovative ability” was included instead of distinguishing every single aspect behind it. In fact, by its very nature, the scientific approach encompasses problem-solving skills and critical thinking, which are commonly found in both STEM and CS activity design.

Table 3 provides a structured set of features that serve as a foundation for the analytical approach adopted in this study. The determined features collectively constitute a structured foundation that guides and informs the analytical methodologies presented in Section 5. Each feature can inform survey construction, ensuring that relevant dimensions are explicitly investigated, or can be used as a factor in analyzing differences across camp implementations.

4.1. Main program characteristics

Table 4 presents the details of the main characteristics for the design of a CS camp to counteract the gender gap. “Main program characteristics” mainly include features related to the composition of the class, the program type, and main topics, and the presence of specific activities to motivate, engage and guide the participants’ future intentions. As stated, these features guide survey construction and analysis.

“Class size”, “Age range”, “Class hours”, and “Program type” are foundational elements that shape extracurricular programs and should be considered both in planning and evaluation. “Class size”, although ideally favoring a low student-to-teacher ratio for effective engagement, may face constraints due to limited resources. A survey question could investigate whether students felt they received adequate individual attention,

Table 4. Main characteristics.

Feature	Description	Analysis suggestion
Class size	Number of students, student/teacher ratio	Include survey questions on satisfaction with student-to-teacher ratio and available support
Age range	K-12 classification	Compare perceived engagement across different age groups
Race and Gender	Demographic composition	Specific question to investigate female and minoritized sense of belonging
Class hours	Hours per lesson, total number of hours	Investigate workload perception and time adequacy
Program type	Weekend programs, summer camp, after-school programs, mixed programs	Compare engagement across different program types
Main topics and related subjects	CS subject	Investigate the influence of multidisciplinary topics on participants' interest in CS
Motivating and engaging context	Motivating and engaging participants through various means	Monitor participant engagement through real-time feedback
Inclusiveness	Attention to inclusion of underrepresented groups	Use survey questions about obstacles faced in joining and feeling included
Positive community	Sense of personal, intellectual, and socio-emotional safety	Investigate students' sense of community using qualitative reflection questions
Support network	Partnership with external networks and NGOs	Assess the perceived impact of external mentors and role models
Agency and choice	Support for student future plans	Measure shifts in career interest and perceived confidence in CS-related decisions

allowing an assessment of whether the student-to-teacher ratio was effective. The "Age range" of participants is a key consideration, as tailored topics and activities significantly vary with different age groups. When designing surveys, this factor may translate into questions assessing whether the content was perceived as age-appropriate and engaging. "Class hours" and "Program type" are interconnected factors; outreach camps, with their condensed time frame, offer the opportunity for extended hours per day over a few days or weeks. This stands in contrast to after-school programs, which may provide fewer hours per week but eventually cover the entire curricular year. To evaluate the impact of these formats, surveys can include questions on workload perception and time adequacy.

While the "Main topic and related subjects" of our exploration will undoubtedly revolve around CS, the inclusive nature of multidisciplinary perspectives allows for related subjects to be centered on one or more specific disciplines. For instance, in the case of computer game development, activities potentially incorporate elements from the realms of arts and engineering alongside CS. Survey questions could explore whether the integration of other disciplines positively influenced participants' interest in CS. Additionally, this could be further investigated by comparing qualitative descriptions of what participants perceive as "computer science" before and after the activity, aiming to assess whether exposure to a multidisciplinary approach has broadened their understanding and influenced their engagement with the field. It is worth noting that while an optimal choice in the main topic and related subjects may automatically create a motivating and engaging context, our experience teaches us that having tutors or assistants in the classroom supporting the main teacher proves invaluable. They play a crucial role in monitoring engagement and facilitate the ability to plan and implement changes effectively to address any emerging issues.

"Race and gender" stand out as key characteristics in our research, with a particular emphasis on addressing gender gaps. While our endeavor involves an effort to ensure adequate representation of women, this initiative aligns closely with the "Inclusiveness" feature. A well-designed evaluation should investigate not only participation rates but also students' experiences in diverse classroom settings. For instance, questions could assess whether female participants felt comfortable engaging in discussions, whether they encountered implicit biases, and how their sense of belonging evolved during the program. Notably, a key aspect of any inclusive initiative should be to target women, individuals from disadvantaged socio-economic backgrounds, from underrepresented racial and ethnic groups, and any other marginalized communities that may vary depending on the socio-political context in which the initiative is situated. Recognizing and addressing the needs of minoritized groups is crucial for ensuring that the initiative is truly inclusive and equitable (Rankin et al., 2020). For example, in line with this goal, the Digital Girls (Faenza, Canali, Carbonaro, et al., 2021c) initiative is offered entirely free of charge and is reserved for female participants. The "Inclusiveness" aspect, while easy to approach in the case of widespread stereotypes, requires organizers to proactively identify and address specific potential issues within each unique setting, i.e. diverse skill levels or accessibility needs. A structured evaluation should thus include an accessibility component, with survey items investigating whether cost, location, or cultural factors acted as barriers to participation. Additionally, to proactively identify the most effective approaches for enhancing accessibility, existing literature should be reviewed where available, and complementary methods such as focus groups and interviews can be employed to gather qualitative insights before finalizing the initiative design. A previous study by the authors (Faenza, et al., 2021b) involved a preliminary screening of some well-known CS extracurricular initiatives in Italy. Upon revisiting the list, it becomes apparent that inclusiveness is not a significant feature of these initiatives. While a limited number of initiatives were free and explicitly addressed women, the geographical distribution of such initiatives reflects the disparity in development across regions, as observed in the previous section in the work by Zheng et al. (2022). To better assess these disparities, evaluation strategies should incorporate geospatial analysis of program availability, as well as participant surveys examining perceived accessibility and potential obstacles to enrollment. These observations highlight the need for policymakers to promote initiatives that address inclusiveness in CS extracurricular activities. The social and family contexts surrounding students must be carefully considered. To evaluate this factor, surveys and interviews should explore parental attitudes toward computing, the extent of family support and students' prior exposure to CS-related activities at home or in their communities. Additionally, structured reflection activities within the camp itself could provide insights into how students' perceptions of CS evolve over time and whether the program fosters a more positive emotional connection to the field.

The "Positive community" feature is also closely linked to inclusiveness, aiming to create an environment where students feel emotionally and intellectually safe. Surveys can explore students' perceptions of the classroom climate, their comfort in asking questions, and whether they experienced stereotype threats. An effective way to assess the impact of a supportive community is to analyze how participants' perceptions of computer science professionals evolve over time. For example, a study such as Semmens et al. (2015) asked participants to list attributes to describe a CS professional and themselves before and after an intervention, and then noticed a reduction in stereotypical and an increase in positivity of those attributes after the initiative.

It is important to note that a positive community will foster a robust network, creating an environment where students feel secure in seeking suggestions and guidance. This open channel for communication not only supports immediate learning but also lays the foundation for students to reach out for future career and educational counseling.

"Support network" refers, in fact, to the collaboration with external networks and NGOs that can provide additional resources, mentorship, and opportunities for students. These partnerships are crucial for expanding the impact of the program beyond the classroom and ensuring students have access to ongoing support. For instance, inviting professionals from relevant industries to speak or mentor can help students see real-world applications of their learning and inspire them to pursue careers in CS. Additionally, this kind of partnership can help secure funding for the program. Evaluation should consider whether these initiatives positively influenced participants' aspirations.

"Agency and choice" involves empowering students to make informed decisions about their future paths by providing them with counseling and mentorship opportunities. This feature is essential for helping students connect their experiences in the program with their long-term educational and career goals. Programs should ensure that students have the opportunity to remain in contact with mentors they identify with if they choose to do so. Surveys should investigate whether participants gained a clearer understanding of CS career paths and whether they felt supported in exploring further learning opportunities.

4.2. Objectives

"Teamwork and communication" enable participants to leverage their unique skills and expertise to collaborate and support each other. The fundamental principle remains constant in different implementations of teamwork activities: students should perceive the advantage of teamwork. This entails working on tasks where they recognize the inherent challenge as attainable only through collective participation (Thibaut et al., 2018).

Surveys can assess whether participants value collaboration more after the experience and whether they feel they contributed meaningfully to the group. Additionally, observations of group dynamics during activities can help identify engagement levels and potential barriers to collaboration.

One of the most crucial features among those presented in Table 5 is the "Informal learning" aspect. Despite the abundance of technological tools available for diverse learning experiences, traditional lecture-style teaching methods still dominate many educational settings. On the other hand, assessing whether student-centered learning

Table 5. Objectives.

Feature	Description	Analysis and Survey Suggestions
Teamwork and communication	Students are encouraged to collaborate actively	Surveys teamwork perception, measure engagement in group tasks, and analyze team roles via observations
Informal learning	Learning throughout experiences not commonly associated with teacher-centered classrooms	Investigate engagement levels in non-traditional learning settings
Scientific spirit and innovative ability	Approach to problems in a scientific and innovative way, learning problem-solving and critical thinking skills	Survey participant confidence in tackling open-ended challenges

leads to better engagement and knowledge retention is essential. In the case of Computational Thinking, unconventional tools such as board games or video games could prove a convenient solution to the matter (Kuo & Hsu, 2020; Leonard et al., 2016).

Lastly, given the nature of most CS-related activities, problem-solving and critical thinking are fundamental skills necessary for success. However, program designers should not assume these skills are innate to computer science-related activities, hence the presence of the "Scientific spirit and innovative ability" feature. To measure their development, participants can be asked to complete problem-solving tasks before and after an intervention, and structured self-reflections can help assess whether they perceive growth in their critical thinking abilities.

4.3. Knowledge content

Even though the knowledge category's main feature, "CS focused", might seem evident, defining explicit content within the vast spectrum of CS to teach is of paramount importance. For effective evaluation, it is necessary to determine whether participants gained the intended CS knowledge. This can be achieved through pre- and post-assessments, tracking conceptual understanding, and skill acquisition.

The "Multidisciplinary" feature in Table 6 advocates for the incorporation of various disciplines beyond CS. This may manifest as a pure multidisciplinary approach, involving the teaching of distinct disciplines and leaving it to the students to establish potential connections between the different content areas. To assess its impact, surveys can include questions measuring how participants perceive CS in relation to other disciplines before and after participation. Additionally, qualitative reflections can capture whether exposure to a multidisciplinary approach influenced their interest in CS. Similarly, "Connection to real-world problems" is a valuable feature in CS education. For instance, designing a program activity centered on tackling climate change naturally requires a multidisciplinary approach. To analyze its effect, surveys can explore whether participants see CS as more relevant to their lives and career aspirations after engaging in problem-solving activities related to societal challenges. For example, in a recent hackathon organized by our group, we provided participants with a series of challenges related to optimizing greenhouse projects among contemporary energy and water crises. This hackathon's outcomes were primarily technological, but other scientific and economic principles were also involved and, in most cases, self-explored by participants.

In the context of developing effective CS programs, the selection of appropriate activity content is of paramount importance in order to employ a "Student-centered

Table 6. Knowledge content.

Feature	Description	Analysis and Survey Suggestions
CS focused	CS subjects are well integrated into activities	Survey participants' confidence in applying new concepts
Multidisciplinarity	Other disciplines outside of CS and STEM are involved	Survey changes in perception of CS as a multidisciplinary field
Connection to real-world problems	Proposed problem originates from real-world	Measure participants' perception of CS relevance before and after
Student-centered approach	Lessons and activities are calibrated on student background knowledge	Compare the effectiveness of different approaches

approach". While it may be tempting to focus on teaching a programming language, selecting the specific language can be daunting. This challenge is intensified in the case of voluntary extracurricular programs where participants may have varying levels of prior knowledge. For instance, a common dilemma program designers face is whether to teach coding with block-based coding languages. Surveys can track student preference and perceived effectiveness of instructional tools (block-based vs. text-based coding), and structured comparisons across different program iterations can inform best practices.

4.4. Activity design

Table 7 reveals that features associated with activity design are the most commonly shared among the characteristics collected from the literature works, starting from the pursuit of an "Interdisciplinarity approach". An "Interdisciplinary approach" entails the integration of diverse disciplines or fields of knowledge to address complex issues, fostering a holistic understanding. Different from the multidisciplinary approach, the interdisciplinary approach supports an implicit connection to other disciplines by proposing activities that may focus on one or more specific disciplines while drawing on contexts from others to enhance and complete the learning experience. To evaluate the effectiveness of an interdisciplinary approach, open-ended questions in Pre- and Post-Surveys can be used to explore how participants perceive connections between disciplines.

"Engineering design" in education involves applying a systematic problem-solving process inspired by engineering principles. It encourages students to identify challenges, generate creative solutions and iteratively refine their designs, fostering critical thinking, collaboration, and hands-on application of theoretical concepts. Assessing students' ability to engage in this process can involve tracking the evolution of their solutions, analyzing how they approach iteration and surveying their perception of engineering-related problem-solving before and after the activity.

Various "Teaching methodologies", such as problem-based learning (PBL), project-based learning and inquiry-based learning(IBL), readily adapt to computer science education. These methodologies, in fact, are already widely employed in CS education; in LaForce et al. (2017) authors suggest that Problem-Based Learning effectively increases

Table 7. Activity design.

Feature	Description	Analysis and Survey Suggestions
Interdisciplinary approach	Interdisciplinary approach to students' learning	Analyze self-reported engagement with different disciplines
Engineering design	Systematic problem-solving process	Track student design iterations, survey confidence in structured problem-solving before and after, analyze solution refinement over time
Teaching methodology	PBL, Project-based learning, IBL	Analyze project quality and problem-solving depth
Presentation and reflection	Students are asked to present and argue solutions	Conduct qualitative analysis of presentations
STEAM approach	Students are encouraged to broaden creativity and foster both holistic and inclusive thinking by incorporating Arts disciplines.	Survey shifts in students' perceptions of creativity in CS, compare STEM and STEAM outcomes
Informal venue	Learning outside of the classic school venue	Investigate whether informal settings increase engagement

student interest in CS disciplines; at the same time, the author emphasizes the importance of the quality of a PBL activity in achieving this goal. To evaluate their impact, the quality of their project outcomes can be analyzed.

"Presentation and reflection" involves providing students with opportunities to showcase their work and reflect on their learning process. This feature encourages students to articulate their cognitive processes, defend their solutions, and receive feedback, which is crucial for developing critical thinking and communication skills. A final presentation moment, where students present their projects and discuss the challenges they faced and the solutions they developed, is highly recommended to reinforce learning and build confidence. Evaluating the impact of these presentations on learning outcomes can include peer assessment, self-reflection surveys, and qualitative analysis of participant discussions.

It is meaningful that in the design of the Digital Girls experience Faenza et al. (2021a), a final presentation moment has consistently been reserved for groups to showcase their work, with emphasis placed on the creative aspect of the activities and the process that led to the final result being presented. In fact, within the main program structure, lessons focused on storytelling and arts have been included to aid participants in developing more enjoyable video games.

The "STEAM (Science, Technology, Engineering, Arts, and Mathematics) approach" encourages students to broaden creativity and foster holistic and inclusive thinking by incorporating Arts disciplines. The "STEAM approach" represents a departure from the traditional focus on better transmitting hard knowledge, shifting the emphasis toward creative processes, integration across disciplines, and inclusive forms of engagement. As highlighted by watson2013transitioning, the shift in emphasis aims at cultivating a higher quality of inclusive thinking. At the same time, the interdisciplinary and creative nature of STEAM initiatives introduces additional challenges for evaluation, as learning outcomes are often diffuse, emergent, and highly context-dependent. Recent work in the field emphasizes that evaluating STEAM activities requires accounting for this increased complexity, particularly when creative and artistic dimensions are interlaced with technical content (Cutrupi et al., 2025). To evaluate the effectiveness of a STEAM approach, Pre- and Post-activity surveys can measure students' perceptions of creativity in STEM before and after participation. Additionally, a qualitative analysis of student projects can assess the integration of artistic and technical elements.

Moreover, the "Informal Venue" feature advocates for learning experiences outside the conventional school setting, removing restrictions imposed by the traditional classroom environment. By moving beyond the confines of the conventional school environment, informal venues foster a more dynamic and flexible educational experience, allowing for greater creativity and engagement. This approach encourages diverse and flexible learning opportunities, fostering a dynamic educational experience beyond the confines of formal schooling. To measure its impact, we can compare engagement levels, attendance, and perceived learning gains between activities held in traditional classrooms and informal settings.

In addition to these considerations, it's worth noting that an ongoing debate surrounds the nature of STEAM in relation to STEM. Recent literature reviews by Aguilera and Ortiz-Revilla (2021) have highlighted that both STEM and STEAM education lack a clear, universally accepted conceptual framework within the scientific and educational community. Despite this lack of conceptual clarity, while the appeal of STEAM education might seem evident, especially in terms of fostering student creativity, it presents a distinctive shift in the focus of STEM. STEAM, in essence, redefines the emphasis of STEM education towards a higher quality of inclusive thinking (Watson & Watson, 2013). This shift transcends the conventional boundaries of individual disciplines, encouraging holistic approaches that incorporate the arts to enrich scientific and technological exploration.

4.5. Tools

Navigating the numerous alternatives in "Software and hardware" tools presents a tough challenge for organizers in the decision-making process. It is important to note that our research group's involvement primarily lies in promoting Open Source software and hardware. With this in mind, it is worth mentioning that the selection of Open Source software, particularly Free and Open Source Software (FOSS), offers additional benefits in terms of inclusivity, specifically from the perspective of socioeconomic differences. FOSS, in fact, may enable participants to continue exploring and experimenting with the tools and resources used during the program without incurring additional costs. To evaluate the effectiveness of tool selection, participant surveys can measure accessibility, usability, and continued engagement with the selected tools post-program. Additionally, qualitative analysis of student feedback and project outcomes can reveal how effectively these tools support learning and creativity.

Regarding the development and evaluation of CS programs, the use of appropriate "Assessment tools" is paramount for the successful iteration of an evidence-based design approach. Additionally, it is crucial to maintain a detailed program log to facilitate the replication process and analyze the choices made throughout the program design.

The feature of "Integrated Technology Use" represents a shift in perception, acknowledging technology not merely as a tool but as an integral component of the learning experience. To analyze the impact of technology integration, surveys and structured interviews can assess whether students perceive technology as a passive tool or as an active learning driver.

"Venue Support" encompasses every aspect related to supporting the physical or virtual space where activities and lessons take place. This includes considerations for optimal setup, accessibility, and resource availability. Participant feedback on venue-related barriers can help organizers optimize learning conditions.

The feature of "Educational Resources" encapsulates a comprehensive array of materials essential for the facilitation of lessons and activities. This includes detailed lesson logs, multimedia resources, and comprehensive manuals and instructions. Tracking the utilization of resources – such as student engagement with manuals, self-reported resource usefulness, and completion of guided activities – can provide insights into their effectiveness.

In our experience, "Teacher-related factors" are of utmost importance. For example, the selection of teachers has been a significant aspect of the program's design. In the case of Digital Girls, consisting of a target audience of female students, having female teachers has been considered a crucial factor in promoting the creation of role models and fostering student identification with successful female figures in the field. Additionally, the program has placed great emphasis on the use of inclusive language. Given that Italian is a language with

Table 8. Tools.

Feature	Description	Analysis and Survey Suggestions
Software and hardware	Software and hardware used for activities	Survey on tool accessibility, usability, and intention of use post-program
Assessment tools	Assess and produce evidence of activities results	Use literature, field observations, focus groups, and student reflections to improve assessment
Integrated technology use	Switch in the perception, technology not only as a tool	Investigate whether participants perceive technology as a core learning tool
Venue support	Every aspect related to venue support to activity and lessons	Gather participant feedback on venue aspects
Educational resources	Lesson logs, multimedia resources, manuals and instructions	Track usage frequency and perceived usefulness of resources
Teacher-related factors	Well-prepared teachers	Analyze classroom interaction for inclusivity

male and female words, teachers are explicitly instructed to use female terms whenever possible. To evaluate the impact of teacher-related factors, structured surveys and interviews with students can measure whether representation and teaching style influenced their perception of CS as an accessible field. Additionally, the qualitative coding of classroom interactions can assess the use of inclusive language and its perceived effects on student engagement.

A summary of these features, along with corresponding analysis and survey suggestions, is provided in Table 8.

5. Tool in practice: an illustrative example of use

The proposed evaluation tool has been applied to analyze the outcomes of the 2022 edition of the Digital Girls Summer Camp, a summer camp activity focused on computer science disciplines reserved for female students aged between 16 and 18. This application serves as an illustrative example of how the tool can be utilized to extract meaningful insights from outreach initiatives.

The goals of this illustrative analysis will be twofold: first, to evaluate the participants' overall satisfaction with the camp experience; second, to investigate the potential impact of the summer camp on the participants' likelihood of pursuing a career in computer science.

To ensure a structured and efficient analysis process, we utilized Jupyter Notebook (Kluyver et al., 2016) as our primary analytical tool. To make these analysis workflows accessible to all team members, including those without programming experience, we deployed NextPyter (Faenza et al., 2024), a collaborative platform that integrates Jupyter Notebook with the Nextcloud document management system, enabling seamless sharing and real-time collaboration on the analysis notebooks. AC1 this environment, the Pandas library (McKinney, 2010) was employed for proficient data management and manipulation, while the Statsmodels library (Seabold & Perktold, 2010) facilitated our statistical analyses. The methodical approach facilitated by notebooks allowed for organized and reproducible research, demonstrating the practical application of the tool. All notebooks used in this analysis are available in the GitLab repository .[7]

5.1. Formal terminology

This analysis defines specific conventions to enhance clarity and facilitate understanding of the variables used. We will use the abbreviation "[t0]" to denote variables associated with data collected during the Pre-Survey phase and "[t1]" to represent those linked to the Post-Survey phase. As stated, researchers may choose to use different survey platforms, which could result in mismatches between the indicator names used in this analysis and those in their own results. Since CSV or Excel/Calc sheets are common export formats for collected participant responses, each column in these files corresponds to the answers to a single question. In the Table 9, we provide a correspondence between the feature names used in this analysis and the questions asked in the survey to assist researchers in integrating their data with the existing notebooks.

We will utilize the abbreviation "[CC]", an abbreviation for "camp characteristics", to indicate variables that relate to the unique attributes and features of the specific camp; the complete list of such is shown in Table 10 and derives from those identified in section 4. While not all indicators are used directly in the analysis, some specific indicators were added post-analysis to compare different camp characteristics that may lead to varying outcomes. These characteristics were identified because we had the opportunity to conduct multiple iterations of our initiative, allowing us to compare different features in the hope of finding those that best align with our objectives. These characteristics must be added to the data source (e.g. the data frame in our notebooks) when applicable. Even if a researcher does not have the opportunity to iterate their camp with different features, the analysis procedure outlined here will still prove useful for identifying indicators of outcomes within the specific context of their analysis.

The proposed feature list consists of a subset of the available questions in the survey, as the open-text responses require different analysis methods than those provided and proposed in the notebooks.

Table 9. Feature description.

Short name	Feature name	Description
At least one parent works in STEM [t0]	Boolean	Whether one of the parental figures works in a STEM field or not
Hours a week playing video games [t0]	Integer	How many hours per week the participant plays video games
Experienced coding before [t0]	Boolean	Whether experienced coding BEFORE the camp start or not
CS path [t0]	Boolean	Whether the participant included a CS path in their future plans or not BEFORE the camp start
CS path [t1]	Boolean	Whether the participant included a CS path in their future plans or not AFTER the camp end
Express creativity [t1]	5-Point Likert	To what extent the participant was able to express her creativity
Made myself [t1]	5-Point Likert	Perceived knowledge mastery during project development
Had fun [t1]	5-Point Likert	Degree of enjoyment experienced during the activity
Teamwork [t1]	5-Point Likert	Evaluation on teamwork
Camp satisfaction [t1]	5-Point Likert	Camp experience overall satisfaction
Camp project satisfaction [t1]	5-Point Likert	Satisfaction over project developed during the camp
Camp team belong [t1]	5-Point Likert	Team belonging perception during project development

Table 10. Feature description.

Feature name	Type	Description
Class size [CC]	Integer	Number of participants to a specific camp, ranging from 14 to 41
STEAM [CC]	Boolean	Whether a specific camp edition also used the STEAM approach or not
Camp length [CC]	Integer	Camp length in number of weeks, ranging from 2 to 3
Block-based coding [CC]	Boolean	Whether a specific camp edition used a block-based coding language or classical text-based coding language
Main teacher is female [CC]	Boolean	Whether the leading teacher of a specific camp was Female or not

Furthermore, it is important to highlight that the specific context in which the initiative was implemented did not provide sufficient diversity in terms of race or ethnicity to include these dimensions in the analysis. Similarly, the variable representing the participants' school of origin was also excluded from the analysis due to its high degree of correlation with the "Experienced coding before [t0]" feature, resulting in an almost perfect overlap.

5.2. Data collection and methodology

The subsequent case study is based on data collected during the 2022 summer edition of the Digital Girls camp. This event, as the name suggests, is reserved for female participants only and was hosted across seven cities – six in the Emilia Romagna region of Italy and one in Lombardia – with a total of 272 participants, all aged in the range of 16–18 years old. Importantly, all participants were accepted without any selection process, and the camp was free for all attendees. The camps exhibited variations in their activities; for example, some camps incorporated block-based coding languages, whereas others employed text-based coding languages. Additionally, the leadership in some camps included female instructors, while others did not.

To provide context for the analysis, it is important to consider the duration and size of these camps. The shortest camp lasted two weeks, while the longest extended to three weeks. The number of participants in each camp ranged from 14 to 41.

Despite these variations, all activities were centered around computer science disciplines, differing primarily in the type of coding language used – either block-based or text-based. A consistent structure was maintained across all camps, which began with a practical approach to introduce fundamental concepts of the discipline, followed by a project development phase where the participants applied the technology they learned to develop a project of their own.

For instance, in some camps, participants learned video game development using a text-based language like Python (PyGame), while others used a block-based language to realize robots or simple Internet of Things (IoT) devices. We summarized the main characteristics of the camp in Table 11.

Data collection was carried out using the proposed evaluation tool, specifically utilizing the LimeSurvey platform with the provided survey template. Participants were associated with unique tokens, which were used consistently across both the Pre- and Post-Surveys. This was achieved through LimeSurvey's option to collect responses without storing the participants' emails, relying solely on the token for matching Pre- and Post-Survey data. While this workflow was implemented manually using LimeSurvey in the present case

Table 11. Camp characteristic.

Camp location	Class size	STEAM	Camp length(week)	Block-based	Female teacher
Bologna	27	N	3	N	N
Cesena	31	N	3	Y	N
Ferrara	16	N	3	N	Y
Mantova(a)	18	Y	2	Y	N
Mantova(b)	25	N	2	N	N
Modena	41	N	3	N	N
Parma	14	N	3	N	Y
Reggio Emilia	29	Y	3	Y	N

N = NO, Y = YES.

study, the same process can be simplified through the deployment of the dedicated evaluation platform described in burchiellaro2025elevateai, which integrates survey generation, participant management, and data export while preserving institutional control over data and privacy.

Once data collection was completed, the first three notebooks of the tool were executed: "01_data_cleaning", "02_outliers_analysis" and '03_validity_check. During the data-cleaning process, the structured nature of the proposed survey facilitated a straightforward workflow. Both Pre- and Post-Survey data were downloaded, including "question codes" and responses in a "coded format" (options available on LimeSurvey's export interface). These files were then placed in the tool's "sources" folder, and the data-cleaning notebook was executed. This notebook connected Pre- and Post-Survey responses using the default "token" field. If an alternative method was used to link responses, as described in the survey section 3, the notebook can be easily adapted. The notebook also provided an analysis of response times to identify potential outliers, leaving it to the analyst to decide whether to exclude them. In our case, we chose to exclude two responses identified as "speeders" – participants who completed the survey unusually quickly. In the final phase of the data cleaning process, the notebook flagged any missing variables in either the Pre- or Post-Survey, allowing for manual adjustments in case of typographical errors or other issues. This step is obviously not necessary if relying on the default token variable, as it is handled automatically by the LimeSurvey platform. The cleaned data were saved as an intermediate result in the "intermediate_results" folder, ready for further analysis.

The outlier analysis notebook identified any remaining outliers in the dataset and performed aggregation for correlated survey questions. In our analysis, three entries were identified as outliers. We reviewed these responses in detail, including the open-text answers, and decided to exclude them based on our assessment. This step is crucial as it requires contextual knowledge and may involve consulting with the camp's instructors to determine if the outlier reflects any observed situations during the activities. During the aggregation phase, several options are provided, with default settings including mean/median aggregation for questions related to attitudes toward CS, principal component analysis for stereotype perceptions and summation for parent influence. However, the data analyst can choose different methods if some questions are adjusted to better align with the societal context of the initiative.

Finally, the validity notebook provided results on Exploratory Factor Analysis and Cronbach's alpha, as detailed in the "Reliability and Validity Testing" section3. These tests were conducted for each group of coherent variables, and any issues identified could be addressed by dropping variables from the analysis or adjusting the survey structure for future iterations.

As a result of this process, we obtained 194 observations that were ready for analysis.

5.3. Exemplary satisfaction analysis

In this subsection, we aim to investigate the factors influencing participants' satisfaction with the camp experience, as measured by the Likert variable denoted as Camp overall satisfaction [t1].

The analysis will showcase the use of regression analysis models, demonstrating how subsamples of participants with different characteristics – specifically those who have already experienced coding versus those who have not – can provide further insights. The description of the results is based on the notebook "satisfaction_analysis", which is derived from an available general notebook called "regression_analysis". This notebook is designed to be copied and customized for each regression analysis you wish to perform. It is important to adapt the predictor variables and outcomes based on the results of the previous phases. For instance, if a variable has been excluded by EFA, it must also be excluded from this analysis.

To achieve this, it is important to use the "camp_characteristic" notebook to add camp characteristics to the dataset. These characteristics are particularly meaningful when analyzing multiple iterations of the initiative with variations to be compared. For an initiative being held for the first time in a single classroom instance, this phase is not necessary, as the characteristics will be the same. This step becomes relevant in subsequent editions or when different classrooms are involved in the same edition, where factors such as the teacher, teaching topics or classroom approach may differ. Any of the described characteristics in Section 4 could vary in such cases. In our analysis, we added the characteristics specified in the table of characteristics, using a table of association between tokens and specific camp locations to perform this task. Another approach could have been to employ different survey instances for each camp and concatenate them only after completing the camp characteristics phase.

The analysis process is then straightforward and follows the structured approach provided by the notebook.

It is important to note that the Likert scale represents an ordered scale where the intervals between points cannot be assumed to have equal magnitudes. Consequently, the analysis is conducted using the ordered logistic regression method based on the available observations.

As stated, the analysis will concentrate on a set of independent variables, each representing distinct participant responses or specific aspects of the camp that the participant encountered. Table 12 presents the results of the ordered logistic regression, where the target variable is the variable "Camp satisfaction [t1]".

The results describe how different variables affect the logarithmic transformation of the chances of moving on to the next level of camp satisfaction for a one-unit change in the relevant independent variable. In other words, this analysis sheds light on the extent and manner in which each variable contributes to the augmentation of overall camp satisfaction. Upon examining the outcomes as depicted in Table 12, certain variables exhibit *p*-values considered highly significant ($p < 0.001$), significant ($p < 0.05$) or

Table 12. Ordered logit: camp overall satisfaction.

Dep. Variable:	Camp Satisfaction [t1]	Log-Likelihood:	-151.38
Model:	OrderedModel	**AIC:**	342.8
Method:	Maximum Likelihood	**BIC:**	408.1
No. Observations:	194		
Df Residuals:	174		
Df Model:	16		

	coef	std err	z	P > |z|	[0.025	0.975]
Class size [CC]	0.0236	0.029	0.811	0.418	−0.033	0.081
STEAM [CC]	−0.9266	0.535	−1.731	0.083*	−1.976	0.123
Camp length [CC]	0.6206	0.656	0.946	0.344	−0.665	1.906
Block-based coding [CC]	1.5540	0.532	2.920	0.004**	0.511	2.597
Main teacher is female [CC]	1.8261	0.794	2.300	0.021**	0.270	3.382
At least one parent works in STEM [t0]	0.2219	0.399	0.557	0.578	−0.559	1.003
Hours a week playing video games [t0]	0.0199	0.016	1.277	0.202	−0.011	0.050
CS path [t0]	0.0564	0.417	0.135	0.893	−0.761	0.874
Experienced coding before [t0]	−0.1962	0.361	−0.543	0.587	−0.904	0.511
Express creativity [t1]	0.6415	0.243	2.638	0.008**	0.165	1.118
Made myself [t1]	0.2026	0.163	1.244	0.213	−0.116	0.522
Had fun [t1]	1.2261	0.253	4.839	0.000***	0.730	1.723
Teamwork [t1]	−0.2744	0.232	−1.184	0.236	−0.729	0.180
CS path [t1]	0.0780	0.440	0.177	0.859	−0.783	0.939
Camp project satisfaction [t1]	1.3259	0.253	5.250	0.000***	0.831	1.821
Camp team belong [t1]	0.1763	0.216	0.815	0.415	−0.248	0.600
1/2	8.1543	1.998	4.082	0.000	4.239	12.070
2/3	1.0967	0.276	3.971	0.000	0.555	1.638
3/4	1.0764	0.142	7.596	0.000	0.799	1.354
4/5	1.4676	0.098	14.986	0.000	1.276	1.660

Note: Marginality of *p*-values is highlighted with the notation: $*p < 0.10$, $**p < 0.05$, $***p < 0.001$.

marginally so ($p < 0.10$), implying evidence against the null hypothesis. In simpler terms, this suggests that the observed outcomes are improbable to arise purely by chance. Notably, the coefficients of these significant variables merit consideration because they encapsulate the extent of alteration in the outcome for a one-unit shift in the respective independent variable, with all other factors held constant.

The coefficients of the variables provide crucial information on their respective influences, where both positive and negative impacts are observed. Notably, the variables "I had fun [t1]" and "Camp project satisfaction [t1]" exhibit substantial positive correlations with camp satisfaction, implying that participants who reported higher levels of carried-out project satisfaction and enjoyment tend to have a more favorable camp experience.

In order to detect possible multicollinearity, we employed the Variance Inflation Factor (VIF) test in our analysis (Midi et al., 2010), considering, as a general rule of thumb, features with a VIF index lower than 4 (O'Brien, 2007) as acceptable. We found that none of the features in our models exceeded this threshold, indicating an absence of multicollinearity concerns.

Surprisingly, camp characteristics do not exhibit significant effects on the results, except for the positive influence of a female leading teacher (coefficient: 1.8261, $p = 0.021$) and the coding language type (coefficient: 1.5540, $p = 0.004$), which seems to have a positive impact when the activities require the use of a block-based coding language. Although not reaching conventional significance, the adoption of the STEAM approach (coefficient: −0.9266, $p = 0.083$) showed a marginal effect. Given the relevance of this characteristic, we conducted a closer qualitative analysis, examining participants' suggestions and post-camp debriefing observations. This revealed that misunderstandings arose due to expectations of more traditional, curricular-style coding lessons.

While the presented results may already offer valuable insights into the factors influencing camp satisfaction, our practical experience suggests that specific aspects related to participants' backgrounds may represent a significant distinguishing factor. For this reason, we carry out a further analysis based on the students' coding experience. We divide the dataset into two subsets consisting of participants with and without prior coding exposure, respectively 78 participants (40%) and 116 participants (60%).

Upon examining the results obtained by analyzing the subset, shown in Table 13, certain variables, specifically "Camp project satisfaction t1" and "Had fun t1", continue to exhibit significant positive associations with camp satisfaction, aligning with the broader analysis.

Regarding camp characteristics, the presence of a female leading teacher and block-based coding language adoption appear to have a reduced impact on participants with coding experience (respectively $p = 0.031$ and $p = 0.025$) if compared to those without (respectively $p = 0.072$ and $p = 0.084$). This result may be attributed to a pre-existing male role model corresponding to a previous teacher figure in the CS field.

Intriguingly, while the results for the other variables for the subset without previous coding experience show marginal significance ($p < 0.10$), the small magnitude of the margin suggests the need for additional investigation to understand their potential impact better. Specifically, "Express creativity [t1]" features hold a positive impact, while "Teamwork[t1]" shows a tiny negative impact. The same negative impact is also noted in the analysis of other outcomes, such as future intentions, leading us to investigate open-ended questions and collect feedback from teachers. While the specific analyses of other outcomes are not included here, as they fall outside the scope of this study, the methodology that led us to draw some conclusions is shared for guidance purposes.

An important aspect to highlight is the internal organization of the initiative. In fact, participants autonomously engaged in project development during the second part of the camp, selecting their roles within the project, with instructors acting as facilitators rather than directors of tasks or roles. Based on teacher feedback, we observed that girls

Table 13. Ordered logit: camp overall liking with coding experience subsets.

	Coding Experience			No Coding Experience		
Variable	Coef.	Std. Err.	P > |z|	Coef.	Std. Err.	P > |z|
Class size [CC]	0.1010	0.061	0.100	0.0198	0.038	0.600
STEAM [CC]	−1.1732	0.902	0.193	−0.6582	0.750	0.380
Camp length [CC]	1.3683	1.287	0.288	−0.0633	0.853	0.941
Block-based coding [CC]	1.5505	0.896	0.084*	1.7869	0.797	0.025**
Main teacher is female [CC]	2.6773	1.487	0.072*	2.3859	0.108	0.031**
At least one parent works in STEM [t0]	0.3393	0.664	0.609	0.2090	0.537	0.697
Hours a week playing video games [t0]	0.0080	0.024	0.734	0.0169	0.024	0.479
CS path [t0]	−0.3096	0.864	0.720	0.3447	0.528	0.513
Express creativity [t1]	0.7063	0.441	0.109	0.5803	0.315	0.065*
Made myself [t1]	0.2829	0.274	0.302	0.1657	0.223	0.458
Had fun [t1]	0.9690	0.432	0.025**	1.5168	0.345	0.000***
Teamwork [t1]	−0.0047	0.394	0.990	−0.5437	0.307	0.077*
CS path [t1]	0.5027	0.956	0.599	−0.2393	0.553	0.665
Camp project satisfaction [t1]	2.0233	0.457	0.000***	1.1327	0.357	0.001**
Camp team belong [t1]	−0.3036	0.415	0.465	0.4460	0.270	0.099*

Note: Marginality of p-values is highlighted with the notation: *$p < 0.10$, **$p < 0.05$, ***$p < 0.001$.

from diverse backgrounds were able to integrate their interdisciplinary knowledge effectively. However, observations during the activities led to the conclusion that the nature of some activities, like video game development, requires various complementary tasks such as storytelling, planning, graphical design, and other non-coding activities. This dynamic may have inadvertently led to some participants engaging less in coding activities or not coding at all during the project phase, potentially influencing the results related to coding interest. An in-depth analysis, perhaps through focus groups, could shed further light on this issue. For the next iteration, we plan to adopt a dual approach: selecting an activity that intrinsically balances coding and non-coding tasks while also incorporating a specific question about the roles participants undertook during the project phase for future analysis.

6. Discussion

This paper aims to synthesize and present a comprehensive evaluation tool designed to assess the effectiveness of educational initiatives, particularly those focused on encouraging female participation in computer science. The tool was developed through a rigorous design research methodology (Cobb et al., 2015), incorporating both qualitative and quantitative approaches to ensure its adaptability across various contexts and iterations of similar educational programs.

The tool's application was demonstrated through a case study, which illustrated how it can be used to evaluate key outcomes such as participant satisfaction and the likelihood of pursuing a future in CS. The case study was an illustration of the tool's practical application and a demonstration of its usability and adaptability in real-world settings. By utilizing regression analysis models, the tool was able to identify factors that influence outcomes, demonstrating its capacity to provide actionable insights for the design and improvement of educational initiatives.

A key strength of the tool lies in its flexibility, allowing it to be tailored to different contexts and participant groups. This flexibility is further supported by the availability of a ready-to-use, deployable platform (Burchiellaro et al., 2025) that simplifies the practical adoption of the evaluation workflow while preserving methodological rigor. With all the notebooks used being available, researchers can both draw from existing analyses or adapt them to their specific needs. For instance, the analysis section highlighted the importance of considering prior experiences, such as coding exposure, when assessing outcomes. In our specific case, the societal context does not provide us with enough racial diversity, but adapting the analysis to take into account this aspect in other contexts would be as straightforward as the proposed masking of coding experience.

The process of data cleaning and preparation was a critical component of the tool, ensuring the reliability and validity of the results Knekta et al. (2019). By systematically linking Pre- and Post-Survey responses and addressing outliers, the framework provides a robust foundation for conducting thorough and accurate evaluations, particularly in the case of adaptation. In fact, it is important to remember that the tool may require further validation when applied in different cultural or educational contexts. The findings generated through its use in one context may not necessarily generalize to others without appropriate adjustments. The notebooks guide the researcher in verifying any changes and adaptations to the proposed survey, fostering a scientific approach to the matter.

The emphasis on using a combination of notebooks for data analysis and data preparation processes, along with the provided instructions to employ a ready-to-use survey with an Open Source platform, makes the tool a valuable tool for educators and researchers who need a clear path for understanding the impact of various educational strategies and for making data-driven decisions to enhance program effectiveness.

In conclusion, the evaluation tool presented in this paper offers a robust and adaptable tool for assessing the impact of CS outreach initiatives. Its application through a case study demonstrated its practical utility and highlighted the importance of considering contextual factors in the evaluation process. As the tool continues to evolve, it holds significant potential for contributing to the development of more effective and inclusive CS education programs. Future research should focus on further refining the tool and exploring its application across different contexts to fully realize its potential as a standard tool for educational evaluation.

7. Limitations

The proposed tool is specifically designed for use in outreach camps, where the primary aim is to encourage interest in computer science among young participants. We are also testing the same tool in the broader STEM field through recent outreach activities led by our research group, adapting the tool with minor modifications to suit the more general STEM context.

While longitudinal studies are ideal for assessing the long-term impact of these initiatives, the inherent nature of outreach programs often presents challenges in collecting long-term data. This is especially true in many European countries, where tracking participants' university intentions can be difficult due to privacy laws and educational system structures. Therefore, we rely, as stated, on Bandura's theory Bandura and Wessels (1997) for assessing self-efficacy, while for assessing future intentions, a crucial outcome for outreach programs focused on STEM or CS, we draw on Ajzen's Theory of Planned Behavior Ajzen (2001).

Given that outreach programs often occur annually or more frequently, the need for iterative data collection is imperative. Such frequent data allow for adjustments and improvements to be made in real-time rather than waiting for the completion of longitudinal studies. Obviously, introducing a third iteration to the evaluation tool, in which intentions recorded earlier are later confirmed or refuted, would enhance our understanding of the program's impact.

8. Conclusions

Our research contributes to the existing body of literature on CS education activities by providing an evaluation tool that defines a replicable procedure for assessing the impact of outreach initiatives on female students' future study and career plans. Specifically, the proposed evaluation tool can be used to identify the strengths and weaknesses of different teaching methods and activities, with the ultimate goal of supporting teachers and researchers in refining and planning future educational initiatives. This contribution is further strengthened by the availability of a ready-to-use, deployable platform that lowers the practical barriers to adopting the proposed evaluation procedure.

In this paper, we described the evaluation tool in its main parts, along with technical suggestions on the practical use of the tool itself. Moreover, we present a case study analysis of the results achieved by applying the tool to the 2022 edition of the Italian Digital Girls summer camp. On one side, the carried-out analysis aims to demonstrate the validity of the tool and its capability to identify the key factors that impact participants' satisfaction and future plans. On the other hand, the evaluation section represents an example of how the collected data can be exploited to carry out the analysis. The complete list of detailed questions used in the evaluation tool and an importable version of the survey designed for LimeSurvey has been made available at a public GitLab link,[8] as stated in Section 3. Additionally, a complete set of ready-to-use notebooks is available in the repository, and these tools are likely to evolve through their use. We hope to foster a community around this project, encouraging contributions and collaboration to improve and expand the tool continuously. When appropriate, the evaluation workflow can also be supported by the associated platform, which integrates survey deployment and data preparation while preserving institutional control over data and privacy.

Beyond offering a structured evaluation framework, this open repository fosters a collaborative community of contributors interested in expanding the tool's capabilities. Researchers and practitioners can propose improvements, add new analysis scenarios, including qualitative assessments, and refine documentation to enhance usability for a broader audience. At the same time, making our research efforts readily available ensures that the tool can be continuously refined and adapted, providing a dynamic and evolving resource for the CS education community.

Acknowledgment

This work was supported by the University of Modena and Reggio Emilia – Fondazione di Modena Project "NextPyter: Open-Source Collaborative Platform for Interdisciplinary Research" funded by Fondo di Ateneo per la ricerca Anno 2024 – Bando per il finanziamento di progetti di ricerca interdisciplinari.

Notes

1. https://research-and-innovation.ec.europa.eu/knowledge-publications-tools-and-data/publications/all-publications/she-figures-2021_en
2. https://www.ifoa.it/servizi/enti-scuole-e-internazionale/progetti-internazionali/stem-for-future/
3. https://gitlab.com/frfaenza/stem-camp-evaluation
4. http://foss2serve.org/index.php/Evaluation_Instruments
5. https://www.limesurvey.org/
6. https://gitlab.com/frfaenza/stem-camp-evaluation
7. https://gitlab.com/frfaenza/stem-camp-evaluation
8. https://gitlab.com/frfaenza/stem-camp-evaluation

Disclosure statement

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

ORCID

Francesco Faenza http://orcid.org/0000-0002-7258-7192
Riccardo Mescoli http://orcid.org/0009-0005-3115-2378
Linda Burchiellaro http://orcid.org/0009-0008-1098-1398
Claudia Canali http://orcid.org/0000-0001-8448-7693

References

Aguilera, D., & Ortiz-Revilla, J. (2021). Stem vs. steam education and student creativity: A systematic literature review. *Education Sciences*, *11*(7), 331. https://doi.org/10.3390/educsci11070331

Aivaloglou, E., & Hermans, F. (2019). Early programming education and career orientation: The effects of gender, self-efficacy, motivation and stereotypes. *Proceedings of the 50th ACM technical symposium on computer science education* (pp. 679–685 doi:https://doi.org/10.1145/3287324.3287358).

Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology*, *52*(1), 27–58. https://doi.org/10.1146/annurev.psych.52.1.27

Archer, L., Dawson, E., DeWitt, J., Seakins, A., & Wong, B. (2015). "Science capital" : A conceptual, methodological, and empirical argument for extending bourdieusian notions of capital beyond the arts. *Journal of Research in Science Teaching*, *52*(7), 922–948. https://doi.org/10.1002/tea.21227

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A., & Wessels, S. (1997). *Self-efficacy*. Cambridge University Press Cambridge.

Binns, I. C., Polly, D., Conrad, J., & Algozzine, B. (2016). Student perceptions of a summer ventures in science and mathematics camp experience. *School Science and Mathematics*, *116*(8), 420–429. https://doi.org/10.1111/ssm.12196

Braswell, K. M., Johnson, J., Brown, B., & Payton, J. (2021). Pivoting during a pandemic: Designing a virtual summer camp to increase confidence of Black and Latina girls. *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* (pp. 686–691).https://doi.org/10.1145/3408877.3432380.

Burchiellaro, L., Faenza, F., & Canali, C. (2025). Elevate-AI: Evaluation of learning environments via assessment tools enhanced by AI. *20th Conference on Computer Science and Intelligence Systems (FedCSIS)*. https://doi.org/10.15439/2025F3908.

Burge, J. E., Gannod, G. C., Doyle, M., & Davis, K. C. (2013). Girls on the go: A CS summer camp to attract and inspire female high school students. *Proceeding of the 44th ACM technical symposium on Computer science education* (pp. 615–620). https://doi.org/10.1145/2445196.2445376.

Cobb, P., Jackson, K., & Dunlap, C. (2015). Design research: An analysis and critique. In *Handbook of international research in mathematics education* (pp. 481–503).Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9780203448946-24/design-research-paul-cobb-kara-jackson-charlotte-dunlap.

Craig, A. (2016). Theorising about gender and computing interventions through an evaluation framework. *Information Systems Journal*, *26*(6), 585–611. https://doi.org/10.1111/isj.12072

Cutrupi, C. M., Diaz, P., Faenza, F., Jaccheri, L., Canali, C., Aedo, I., Kouluontzos, V., & Galani, C. (2025). Creative and arts-based practices to engage young girls in STEM-preliminary insights from a qualitative study. *2025 World Engineering Education Forum-Global Engineering Deans Council (WEEF-GEDC)* (pp. 1–9).IEEE. https://doi.org/10.1109/WEEF-GEDC66748.2025.11256327.

Danoff, M. (2017). *Gender and computer science at Harvard* [PhD thesis]. Harvard College.

Davis, K. E. B., & Hardin, S. E. (2013). Making stem fun: How to organize a STEM camp. *Teaching Exceptional Children*, *45*(4), 60–67. https://doi.org/10.1177/004005991304500408

Decker, A., & McGill, M. M. (2019). A topical review of evaluation instruments for computing education. *Proceedings of the 50th ACM Technical Symposium on Computer Science Education, SIGCSE '19* (pp. 558–564). Association for Computing Machinery, New York, NY, USA.

Decker, A., McGill, M. M., & Settle, A. (2016). Towards a common framework for evaluating computing outreach activities. *Proceedings of the 47th ACM Technical Symposium on Computing Science Education* (pp. 627–632). https://doi.org/10.1145/2839509.2844567.

European Commission. (2022). *Digital economy and society index (DESI)*. https://digital-strategy.ec.europa.eu/en/library/digital-economy-and-society-index-desi-2022

European Commission. (2023). *Proposal for a council recommendation on improving the provision of digital skills in education and training*. https://education.ec.europa.eu/document/factsheet-proposal-for-a-council-recommendation-on-improving-the-provision-of-digital-skills-in-education-and-training .

Faenza, F., & Canali, C. (2023). An evaluation tool for extracurricular activities to reduce the gender gap in ICT. *International Conference on Gender Research (ICGR)*, Londonderry Northen Ireland.

Faenza, F., Canali, C., & Carbonaro, A. (2021a). Ict extra-curricular activities: The "digital girls" case study for the development of human capital. In A. Visvizi, O. Troisi, & K. Saeedi (Eds.), *Research and innovation forum 2021* (pp. 193–205). Springer International Publishing.

Faenza, F., Canali, C., Carbonaro, A., et al. (2021b). Evaluating different approaches to closing the gender gap at ICT summer camps in Italy. *4th International Conference on Gender Research, ICGR* (pp. 104–113). https://doi.org/10.34190/IGR.21.051.

Faenza, F., Canali, C., Colajanni, M., & Carbonaro, A. (2021c). The digital girls response to pandemic: Impacts of in presence and online extracurricular activities on girls future academic choices. *Education Sciences*, *11*(11), 715. https://doi.org/10.3390/educsci11110715

Faenza, F., Canali, C., Fregni, L., & Maccaferri, E. (2024). NextPyter: Open-Source Research Collaborative Platform. In Practice and Experience in Advanced Research Computing 2024: Human Powered Computing (pp. 1–9 doi:https://doi.org/10.1145/3626203.3670516).

Franklin, D., Conrad, P., Boe, B., Nilsen, K., Hill, C., Len, M., Dreschler, G., Aldana, G., Almeida-Tanaka, P., Kiefer, B., et al. (2013). Assessment of computer science learning in a Scratch-based outreach program. *Proceeding of the 44th ACM technical symposium on Computer science education* (pp. 371–376). https://doi.org/10.1145/2445196.2445304.

Fronza, I., Corral, L., Pahl, C., & Iaccarino, G. (2020). Evaluating the effectiveness of a coding camp through the analysis of a follow-up project. *Proceedings of the 21st Annual Conference on Information Technology Education, SIGITE '20* (pp. 248–253). Association for Computing Machinery, New York, NY, USA.

Gabay-Egozi, L., Shavit, Y., & Yaish, M. (2015). Gender differences in fields of study: The role of significant others and rational choice motivations. *European Sociological Review*, *31*(3), 284–297. https://doi.org/10.1093/esr/jcu090

Guzey, S. S., Moore, T. J., & Harwell, M. (2016). Building up STEM: An analysis of teacher-developed engineering design-based STEM integration curricular materials. *Journal of Pre-College Engineering Education Research*, *6*(1). Cited by: 88; All Open Access, Gold Open Access, Green Open Access.https://doi.org/10.7771/2157-9288.1129

Heaverlo, C. A., Cooper, R., & Lannan, F. S. (2013). Stem development: Predictors for 6th-12th grade girls' interest and confidence in science and math. *Journal of Women and Minorities in Science and Engineering*, *19*(2), 121–142. https://doi.org/10.1615/JWomenMinorScienEng.2013006464

Jenson, J., de Castell, S., & Fisher, S. (2007). Girls playing games: Rethinking stereotypes. *2007 conference on Future Play* (pp. 9–16). https://doi.org/10.1145/1328202.1328205.

Kallia, M., & Sentance, S. (2018). Are boys more confident than girls? The role of calibration and students' self-efficacy in programming tasks and computer science. *13th Workshop in Primary and Secondary Computing Education*, 1–4). https://doi.org/10.1145/3265757.3265773.

Klein, J. T. (2010). A taxonomy of interdisciplinarity. *The Oxford Handbook of Interdisciplinarity*, *15*(6), 15.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., & Development Team, J. (2016). Jupyter notebooks - a publishing format for reproducible computational workflows. In F. Loizides & B. Scmidt (Eds.), *Positioning and power in academic publishing: Players, agents and agendas* (pp. 87–90). IOS Press.

Knekta, E., Runyon, C., & Eddy, S. (2019). One size doesn"™t fit all: Using factor analysis to gather validity evidence when using surveys in your research. *CBE-Life Sciences Education*, *18*(1), rm1.
Kuo, W.-C., & Hsu, T.-C. (2020). Learning computational thinking without a computer: How computational participation happens in a computational thinking board game. *The Asia-Pacific Education Researcher*, *29*(1), 67–83. https://doi.org/10.1007/s40299-019-00479-9
LaForce, M., Noble, E., & Blackwell, C. (2017). Problem-based learning (PBL) and student interest in STEM careers: The roles of motivation and ability beliefs. *Education Sciences*, *7*(4), 92. https://doi.org/10.3390/educsci7040092
Leonard, J., Buss, A., Gamboa, R., Mitchell, M., Fashola, O. S., Hubert, T., & Almughyirah, S. (2016). Using robotics and game design to enhance children"™s self-efficacy, STEM attitudes, and computational thinking skills. *Journal of Science Education and Technology*, *25*(6), 860–876. https://doi.org/10.1007/s10956-016-9628-2
Lewis, C. M., Anderson, R. E., & Yasuhara, K. (2016). "I don't code all day" fitting in computer science when the stereotypes don't fit. *2016 ACM conference on international computing education research* (pp. 23–32). https://doi.org/10.1145/2960310.2960332.
Lewis, C. M., Yasuhara, K., & Anderson, R. E. (2011). Deciding to major in computer science: A grounded theory of students' self-assessment of ability. *7th International Workshop on Computing Education Research*, 3–10). https://doi.org/10.1145/2016911.2016915.
Li, Y., Wang, K., Xiao, Y., & Froyd, J. E. (2020). Research and trends in STEM education: A systematic review of journal publications. *International Journal of STEM Education*, *7*(1). https://doi.org/10.1186/s40594-020-00207-6
Lynch, S. J., Burton, E. P., Behrend, T., House, A., Ford, M., Spillane, N., Matray, S., Han, E., & Means, B. (2018). Understanding inclusive STEM high schools as opportunity structures for underrepresented students: Critical components. *Journal of Research in Science Teaching*, *55*(5), 712–748. https://doi.org/10.1002/tea.21437
Martn-Páez, T., Aguilera, D., Perales-Palacios, F. J., & Vlchez-González, J. M. (2019). What are we talking about when we talk about STEM education? A review of literature. *Science Education*, *103*(4), 799–822.
McGill, M. M., & Decker, A. (2020). A gap analysis of statistical data reporting in K-12 computing education research: Recommendations for improvement. *Proceedings of the 51st ACM Technical Symposium on Computer Science Education, SIGCSE '20* (pp. 591–597). Association for Computing Machinery, New York, NY, USA.
McKinney, W. (2010). Data structures for statistical computing in Python. In STÉFAN. van der Walt & JARROD. Millman (Eds.). *Proceedings of the 9th Python in Science Conference* (pp. 56–61). https://pdfs.semanticscholar.org/ef4e/f7f38bb907e5d7b4df3e6ff1db269d4970f5.pdf .
Midi, H., Sarkar, S. K., & Rana, S. (2010). Collinearity diagnostics of binary logistic regression model. *Journal of Interdisciplinary Mathematics*, *13*(3), 253–267. https://doi.org/10.1080/09720502.2010.10700699
Mohr-Schroeder, M. J., Jackson, C., Miller, M., Walcott, B., Little, D. L., Speler, L., Schooler, W., & Schroeder, D. C. (2014). Developing middle school students' interests in STEM via summer learning experiences: S ee b lue STEM c amp. *School Science and Mathematics*, *114*(6), 291–301. https://doi.org/10.1111/ssm.12079
National Science Foundation. (2021). Women, minorities, and persons with disabilities in science and engineering. *National Center for Science and Engineering Statistics*. Special Report NSF. https://ncses.nsf.gov/pubs/nsf21321 .
Nicolescu, B. (2014). Multidisciplinarity, interdisciplinarity, indisciplinarity, and transdisciplinarity: Similarities and differences. *RCC Perspectives*, 19–26. https://www.jstor.org/stable/26241230.
Nite, S., Capraro, M., Capraro, R., & Bicer, A. (2017). Explicating the characteristics of STEM teaching and learning: A metasynthesis. *Journal of STEM Teacher Education*, *52*(1). https://doi.org/10.30707/JSTE52.1Nite
O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, *41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6
OPSI. (2023). *Observatory for public sector innovation*. https://oecd-opsi.org/innovations/digital-girls-emilia-romagna/

Rankin, Y. A., Thomas, J. O., & Joseph, N. M. (2020). Intersectionality in HCI: Lost in translation. *Interactions*, *27*(5), 68–71. https://doi.org/10.1145/3416498

Reimann, P. (2010). Design-based research. In *Methodological choice and design: Scholarship, policy and practice in social and educational research* (pp. 37–50).Springer. https://doi.org/10.1007/978-90-481-8933-5_3.

Rorrer, A. S. (2016). An evaluation capacity building toolkit for principal investigators of undergraduate research experiences: A demonstration of transforming theory into practice. *Evaluation and Program Planning*, *55*, 103–111. https://doi.org/10.1016/j.evalprogplan.2015.12.006

Rosson, M. B., Carroll, J. M., & Sinha, H. (2011). Orientation of undergraduates toward careers in the computer and information sciences: Gender, self-efficacy and social support. *ACM Transactions on Computing Education (TOCE)*, *11*(3), 1–23. https://doi.org/10.1145/2037276.2037278

Sax, L. J., Lehman, K. J., Jacobs, J. A., Kanny, M. A., Lim, G., Monje-Paulson, L., & Zimmerman, H. B. (2017). Anatomy of an enduring gender gap: The evolution of women"™s participation in computer science. *Journal of Higher Education*, *88*(2), 258–293. https://doi.org/10.1080/00221546.2016.1257306

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. *9th Python in Science Conference*. https://pdfs.semanticscholar.org/3a27/6417e5350e29cb6bf04ea5a4785601d5a215.pdf .

Semmens, R., Piech, C., & Friend, M. (2015). Who are you? We really wanna know... especially if you think you're like a computer scientist. *Proceedings of the Third Conference on GenderIT* (pp. 40–43). https://doi.org/10.1145/2807565.2807711.

Spieler, B., Oates-Indruchovà, L., & Slany, W. (2020). Female students in computer science education: Understanding stereotypes, negative impacts, and positive motivation. *Journal of Women and Minorities in Science and Engineering*, *26*(5), 473–510. https://doi.org/10.1615/JWomenMinorScienEng.2020028567

Stanko, T., & Zhirosh, O. (2017). Young women who choose it: What role do their families play? *2017 7th World Engineering Education Forum (WEEF)* (pp. 88–93). IEEE.https://doi.org/10.1109/WEEF.2017.8467169.

Thibaut, L., Ceuppens, S., De Loof, H., De Meester, J., Goovaerts, L., Struyf, A., Boeve de Pauw, J., Dehaene, W., Deprez, J., De Cock, M., Hellinckx, L., Knipprath, H., Langie, G., Struyven, K., Van de Velde, D., Van Petegem, P., & Depaepe, F. (2018). Integrated STEM education: A systematic review of instructional practices in secondary education. *European Journal of STEM Education*, *3*(1), 2. https://doi.org/10.20897/ejsteme/85525

Vrieler, T., Nylén, A., & Cajander, Å. (2021). Computer science club for girls and boys-a survey study on gender differences. *Computer Science Education*, *31*(4), 431–461. https://doi.org/10.1080/08993408.2020.1832412

Watson, A. D., & Watson, G. H. (2013). Transitioning stem to steam: Reformation of engineering education. *Journal for Quality and Participation*, *36*(3), 1–5.

Wilson, D., Allendoerfer, C., Kim, M. J., Burpee, E., Bates, R. A., Smith, T. F., Plett, M., & Veilleux, N. M. (2013). Stem students outside the classroom: The role of the institution in defining extracurricular activity. *2013 ASEE Annual Conference & Exposition* (pp. 23–1085. https://peer.asee.org/22470.pdf.

Yasar, O., & Landau, R. H. (2003). Elements of computational science and engineering education. *SIAM Review*, *45*(4), 787–805. https://doi.org/10.1137/S0036144502408075

Yilmaz, M., Ren, J., Custer, S., & Coleman, J. (2009). Hands-on summer camp to attract K-12 students to engineering fields. *IEEE Transactions on Education*, *53*(1), 144–151.

Zailan, N. A., Bunyamin, M. A. H., Hanri, C., Ibrahim, N. H., Osman, S., Ismail, N., & Azelee, N. W. (2019). Assessment and evaluation of non-formal STEM education programs. *International Journal of Recent Technology and Engineering (IJRTE)*, *7*. https://www.researchgate.net/publication/338392214_Assessment_and_Evaluation_of_Non-Formal_STEM_Education_Programs.

Zheng, Y., Liu, P., Yang, X., Guo, Y., Qiu, X., Jin, X., Luo, X., & Zheng, T. (2022). K""12 science, technology, engineering, and math characteristics and recommendations based on analyses of teaching cases in China. In *Frontiers in psychology* (Vol. 13). Cited by: 0; All Open Access, Gold Open Access, Green Open Access. https://doi.org/10.3389/fpsyg.2022.1010033.