How are primary school computer science curricular reforms contributing to equity? Impact on student learning, perception of the discipline, and gender gaps

Laila El-Hamamsy[1,2]* , Barbara Bruno[3], Catherine Audrin[4], Morgane Chevalier[4], Sunny Avry[2], Jessica Dehler Zufferey[2] and Francesco Mondada[1,2]

Abstract

Background Early exposure to Computer Science (CS) and Computational Thinking (CT) for all is critical to broaden participation and promote equity in the field. But how does the introduction of CS and CT into primary school curricula impact learning, perception, and gaps between groups of students?

Methodology We investigate a CS-curricular reform and teacher Professional Development (PD) programme from an equity standpoint by applying hierarchical regression and structural equation modelling on student learning and perception data from three studies with, respectively, 1384, 2433 and 1644 grade 3–6 students (ages 7–11) and their 83, 142 and 95 teachers.

Results Regarding learning, exposure to CS instruction appears to contribute to closing the performance gap between low-achieving and high-achieving students, as well as pre-existing gender gaps. Despite a lack of direct influence of what was taught on student learning, there is no impact of teachers' demographics or motivation on student learning, with teachers' perception of the CS-PD positively influencing learning. Regarding perception, students perceive CS and its teaching tools (robotics, tablets) positively, and even more so when they perceive a role model close to them as doing CS. Nonetheless, gender differences exist all around with boys perceiving CS more positively than girls despite access to CS education. However, access to CS-education affects boys and girls differently: larger gender gaps are closing (namely those related to robotics), while smaller gaps are increasing (namely those related to CS and tablets).

Conclusion This article highlights how a CS curricular reform impacts learning, perception, and equity and supports the importance of (i) early introductions to CS for all; (ii) preparing teachers to teach CS all the while removing the influence of teacher demographics and motivation on student outcomes; and (iii) having developmentally appropriate activities that signal to all groups of students.

Keywords Computer science, Curricular reform, Elementary education, Learning achievement, Computational thinking, Perception survey, Equity, Gender gaps

*Correspondence:
Laila El-Hamamsy
laila.elhamamsy@epfl.ch
Full list of author information is available at the end of the article

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 2 of 43

Introduction and related work

Introducing computer science and computational thinking for all from an equity perspective

The past decades have seen a growing international consensus regarding the importance of teaching Computer Science (CS) and Computational Thinking (CT) to ensure that students are digitally literate (Webb et al., 2017). Computing is increasingly ubiquitous in today's societies, thus leading to CS being more and more often considered as a subset of STEM education which must be rendered as available to students as mathematics or science education (Guzdial & Morrison, 2016). Introducing CS into formal education is also considered to foster Computational Thinking (CT), an essential skill for everyone in the twenty-first century (Jiang & Wong, 2022; Zhang et al., 2023) which is as important as reading, writing, and arithmetics (Wing, 2006). Teaching CT is not only considered by researchers to benefit STEM-related disciplines (Hurt et al., 2023; Swaid, 2015), but is also considered transversal. The benefits of CT are thought to go beyond CS or mathematics (Denning & Tedre, 2021; Li et al., 2020; Mannila et al., 2014; Weintrop, 2021; Weintrop et al., 2016; Zhang et al., 2023), extending to arts (Zhang et al., 2023), with new evidence even showing that young students employ CT during free play (Kotsopoulos et al., 2022), thus providing an additional lever to introduce both CS and CT to all. Although studies on K-12 CS education and CT have increased significantly in recent years (Apiola et al., 2023; Bers et al., 2022b; Hsu et al., 2018), introducing CS and CT into curricula has been a challenge internationally.

Ottenbreit-Leftwich and Yadav (2022) recently expressed the importance of a "system-wide implementation of CT" from an equity perspective to ensure that all students are introduced to CT, and not just those of a select number of teachers who choose to teach CT. This is echoed by Bers et al. (2022b) who advocate that exposure to CS and CT should happen in early foundational years (ages 3–8) "from a social equity perspective to prevent stereotypes and ensure [that] all young children receive equal opportunities to develop their digital literacy". Two key points emerge from this discourse and must be addressed to broaden participation and promote equity in these fields:

- Structural barriers are access-related and limit (early) CS and CT experiences for all, but can be addressed through curricular reforms (Ottenbreit-Leftwich & Yadav, 2022).
- Social barriers, often stereotype (and therefore gender) related, arise despite equal access and regardless of socioeconomic status (Wang & Hejazi Moghadam, 2017), but can be addressed through early exposure to mitigate the effects of existing stereotypes (Bers et al., 2022b).

The consequence of social and structural barriers is that disparities are present at multiple levels, including performance (i.e. learning) and attitudes towards CS (i.e. perception). Such disparities ultimately contribute to having under-represented groups in CS and CT-related fields, and must therefore be addressed in order to increase the likelihood that a more diverse and inclusive set of people persist in these fields. In the following sections, we delve into the literature and highlight the disparities that exist due to such barriers, particularly in terms of learning and perception.

The influence of social and structural barriers on learning-related equity

Several studies have shown that unequal access to (high-quality) CS education (Bers et al., 2022b; Wang & Hejazi Moghadam, 2017) contributes to performance gaps. In particular, a recent large-scale analysis of performance with 46,000 students from 14 countries conducted by Karpinski et al. (2021) found that socioeconomic background, and therefore access, was related to persistent gaps in CT performance. Their findings indicated that students from "less advantaged backgrounds had lower levels of computer skills [...], especially in CT" (Karpinski et al., 2021). Unfortunately, regardless of access, several studies have found that boys perform better than girls (El-Hamamsy et al., 2022c; Kong & Lai, 2022b; Polat et al., 2021; Román-González et al., 2017), even in kindergarten (Sullivan & Bers, 2016), due to the existence of stereotypes (see " The influence of social and structural barriers on equity related to the perception of the discipline").

Although access to developmentally appropriate CS & CT education can increase students' skills from a young age (Bers et al., 2014; 2022a, b; Hall & McCormick, 2022; Relkin et al. 2021), several studies suggest that perception of the discipline can also influence performance (Hinckle et al., 2020; Rachmatullah et al., 2022; Sun et al., 2022). Rachmatullah et al. (2022) for instance found that the gender-performance gap was more prevalent in countries where the "socio-cultural context" tends to promote such stereotypes and "influenc[e] gender diversity in the CS field". Their findings are corroborated by Hinckle et al. (2020) who found that student learning was not directly influenced by prior experience, but was mediated by their perception of CS. Numerous studies in higher education have also found that motivational and affective factors influence performance and participation in the field (Lishinski et al., 2022), and that they are influenced by

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 3 of 43

gender and ethnicity (Lishinski et al., 2022; Warner et al., 2022). These studies confirm the importance of:

- developing CS and CT initiatives that broaden participation to all students,
- considering their impact on performance and perception to verify whether the gaps between different groups of participants are decreasing.

The influence of social and structural barriers on equity related to the perception of the discipline

Perception-related biases are considered to contribute to disparities and under-representation in CS for women (Rachmatullah et al., 2022; Wang & Hejazi Moghadam, 2017), and more generally for under-represented minorities (Lishinski et al., 2022; Warner et al., 2022), due to stereotype threat (i.e. conforming to/inducing a stereotype simply because you know it exists). Unfortunately, the developmental literature has found that basic stereotypes develop in children as young as 2–3 years old (Bers et al., 2022b). This is confirmed by multiple studies that identified CS-related stereotypes in young children (e.g. starting 6 years old, Master et al. 2021, and even kindergarten, Sullivan & Bers, 2016). The result is that when students are exposed to negative CS-stereotypes, students in the stereotyped group (e.g. girls in this case) tend to endorse those beliefs (Plante et al., 2013; Vandenberg et al., 2021) which negatively impacts their performance, motivation, and career intentions (Master & Meltzoff, 2020; Plante et al., 2013; Vandenberg et al., 2021). For instance, Cheryan et al. (2013) found that women who were presented non-stereotypical views on computer scientists were more likely to express an interest in majoring in CS. Therefore, students may make early career decisions informed by such stereotypes, contributing to an early gender gap (Wang & Hejazi Moghadam, 2017), and long-term disparities in the fields of CS and engineering (Master et al., 2021).

As gender-related stereotypes are prevalent, it is not surprising that numerous studies find that girls perceive CS more negatively than boys (El-Hamamsy et al., 2023c; Kong et al., 2018; Vandenberg et al., 2021; Witherspoon et al., 2016), contributing to a lower sense of belonging (Cheryan et al., 2013, 2017; Opps & Yadav, 2022; Vandenberg et al., 2021), self-efficacy (Beyer, 2014; Kong et al., 2018; Vandenberg et al., 2021), and interest (Beyer, 2014; Master et al., 2021). Provided the importance of such factors for academic achievement and career decisions (Bandura, 1993; Beyer, 2014; Howard et al., 2021a; Olivier et al., 2019), the consequence is that CS "suffers from the lowest participation of girls than other science, technology, engineering, and mathematics (STEM)

subjects (Cheryan et al., 2017)" (Hinckle et al., 2020; Jiang & Wong, 2022). As prior experience may positively affect attitudes toward CS (Hinckle et al., 2020), researchers have suggested that engaging early in CS-related activities that "signal equally to both girls and boys that they belong and can succeed" (Cheryan et al., 2017) in CS, may increase girls' interest, and ultimately contribute to addressing gender equity in the field (Cheryan et al. 2017; Hinckle et al., 2020; Jiang & Wong, 2022). Therefore, in the rest of the article we refer to perception-related equity as the reduction of the influence of stereotypes around CS & CT that lead to biases between groups of people (namely gender) and may influence their motivation, engagement, participation and persistence in these fields.

How are CS and CT curricular reforms having an impact and contributing to equity in these fields?

Early CS and CT opportunities for all students are essential to address structural and social barriers, broaden CS participation, and promote equity in the field. An increasing number of initiatives have therefore sought to include CS and CT in compulsory K-12 worldwide (Balanskat & Engelhardt, 2015; Bers et al., 2022b; Bocconi et al., 2022; European Union and Education, 2019; Hubwieser et al., 2015; Voogt et al., 2015; Webb et al., 2017). In this context, it is essential to establish how such initiatives affect students (Guskey, 2002). This should extend beyond learning to include perception, and investigate how these dimensions interrelate (Hinckle et al., 2020) to ensure that expanding CS to K-12 "neither exacerbates existing equity gaps in education nor hinders efforts to diversify the field of CS" (Wang & Hejazi Moghadam, 2017). The student-level impact of widespread CS and CT curricular reforms, and professional development (PD) programmes, is however seldom evaluated. "Studies that relate student's learning achievement and teachers' capacity building are still rare in the literature of CT (Mason & Rich, 2019)" (Kong & Lai, 2022a). This is likely due to the difficulties countries face implementing CS & CT reforms, including adequately training a sufficient number of teachers to teach the new concepts (Bocconi et al., 2022; El-Hamamsy et al., 2021b). Difficulties of assessing teachers' mastery of Computational Pedagogical Content Knowledge (Hickmott & Prieto-Rodriguez, 2018), and what is implemented after PD programmes (El-Hamamsy et al., 2022a) also exist, despite their direct influence on student learning (Kong & Lai, 2022a). To the best of our knowledge, only Kong & Lai (2022a) linked 81 teachers' content knowledge with 3226 students' achievement in their evaluation of a PD programme. However, these teachers chose to participate in the PD programme and were required to teach a year-long curriculum. This

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 4 of 43

differs significantly from mandatory curricular reform contexts, where the PD programme is imposed on all teachers, resulting in teachers who implement the pedagogical content to varying degrees, if at all. Since a prerequisite to achieving equity is that CS-related reforms have an impact, the lack of studies evaluating the impact of CS reforms means that there is little insight into whether these reforms are contributing to equity and reducing learning and perception gaps between different groups of students.

Since a "K-12 curriculum is a zero-sum game, where adding a subject means [removing] something" (Ottenbreit-Leftwich & Yadav, 2022), it is essential to establish the effectiveness of implementing CS & CT curricula in formal education. Evaluating a reform's effectiveness is critical given:

- the need to improve corresponding PD programmes and curricula (Hickmott & Prieto-Rodriguez, 2018),
- the objective of sustaining the reform in teachers' practices (Hubers, 2020),
- the importance of alleviating concerns of funding agencies and government bodies regarding the impact of the reform and PD programme on teachers (Hickmott & Prieto-Rodriguez, 2018) and students.

Studies evaluating the impact of reforms are even more pressing since recent findings indicate that teachers are not necessarily convinced that their students are learning as a result of teaching these novel curricula (El-Hamamsy et al., 2023b; Toh, 2016) . Establishing the benefits at the student-level is therefore not only necessary to have a complete evaluation of reforms (Avry et al., 2022; El-Hamamsy et al., 2023b; Guskey, 2000), but is essential if the objective is to promote teachers' decisions to continue to implement a new practice in the long term (Howard et al., 2021b; Klingner et al., 2001).

Problem statement and research questions

The present study therefore looks to contribute to understanding the influence of CS curricular reforms on student learning and perception and determining to what extent they contribute to equity with respect to: (i) gender, i.e. reducing significant differences between boys' and girls' perception and performance; (ii) performance, i.e. reducing significant differences between initially low and high performers; and (iii) self-efficacy, i.e. reducing significant differences between students who have low or high self-efficacy. Please note that although the main focus of the article is on the former equity dimensions, one must not neglect the importance of equity in terms of socioeconomic status (Vandenberg et al., 2021; Wang & Hejazi Moghadam, 2017), a dimension which we did not have access to in the present context.

We propose to address the overarching question of equity in two steps: first investigating whether and how the reform significantly influences perception and learning (impact), and then how the results differ according to student populations (equity). To that effect, we investigate the impact of a mandatory CS curricular reform and teacher PD programme (see "Context: a computer science curricular reform for all to promote equity starting primary school") to understand *whether and how the primary school Computer Science curricular reform is contributing to reaching equity goals* (i.e. broadening participation in the field to a larger number and a more diverse set of people). We therefore consider the following research questions:

(RQ1) How does teaching CS pedagogical content[1] impact student learning? And how does it impact learning-related gender- and performance-equity?

(RQ2) How does teaching CS pedagogical content impact students' perception of CS and the tools used to teach it (i.e. robots and tablets)? And how does it impact perception-related self-efficacy and gender-equity?

To answer these questions, we employ data collected between January 2021 and June 2022 in the context of a mandatory primary school CS-curricular reform that is presently being deployed to all grade 1–6 teachers in the region after a piloting phase. The data stem from three studies (see Table 1), the first on student learning (RQ1), the second on perception of the discipline and performance (RQ1, RQ2), and the third on perception of the discipline (RQ2). These studies involved, respectively, $n_1 = 1384$, $n_2 = 2433$ and $n_3 = 1644$ grade 3–6 students (ages 7–11) and their $n_1 = 83$, $n_2 = 142$ and $n_3 = 95$ teachers. The data are analysed through hierarchical linear modelling for student learning, and Structural Equation Modelling for perception, to establish the link between teaching CS and these key outcome variables.

Context: a computer science curricular reform for all to promote equity starting at primary school

The research is part of a large-scale project seeking to introduce Digital Education (also referred to as Computing Education) as a new discipline for all students in

---

[1] CS pedagogical content refers to pedagogical activities that intend to teach students about CS, as opposed to those that employ CS as a tool to teach in other disciplines.

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 5 of 43

Table 1 Synthesis of the three studies evaluating the impact of the CS-curricular reform at the student-level

	Study 1: learning	Study 2: perception and performance	Study 3: perception
Date	January and June 2021	November 2021	June 2022
Grades	3–4 (ages 7–9)	3–6 (ages 7–11)	3–6 (ages 7–11)
Number of schools	7 CS-schools	7 CS-schools	3 CS-schools and 2 non-CS-schools
Number of teachers	83	142	95
Number of students	1384	2433	1644
Student—CT-concepts	x	x	
Student—perception of CS		x	x
Teacher—perception of CS	x		
Teacher—activities taught		x	

the Canton of Vaud in Switzerland (El-Hamamsy et al., 2021b). The curricular reform relies on the collaboration between four institutions in the region (the department of education, the university of teacher education, a higher education university and the technical university) within a research practice partnership to develop the curriculum and corresponding mandatory teacher-PD programme for CS, Information and Communication Technology and Digital Citizenship. To ensure the sustainability and scalability of the reform, the project began with a piloting phase with 10 representative schools from the region (hereby referred to as CS-schools) before large-scale deployment. The CS-curriculum and teacher PD-programme was piloted for the first time and iteratively adjusted for grades 1–4 in 2018–2019, and for grades 5–6 in 2019–2020, with all the teachers from the 10 CS-schools (approximately $n_{grades1-4} = 350$, and $n_{grades5-6} = 180$)[2]. This resulted in a reference manual[3] containing pedagogical activities (for CS, $n_{grades1-4} = 13$, $n_{grades5-6} = 12$) that the teachers can choose from to achieve the curricular objectives (in terms of algorithms and languages, machines and networks, information and data, and the impact of CS on society). The teachers were trained to teach these activities during a mandatory CS-PD that they participated in prior to the present study and were encouraged to teach the novel discipline which is now part of the regional study plan. They were however not required to do so. Given that in primary school there is no dedicated hour in the grid for Digital Education (and thus CS), and that the discipline is not evaluated, this leads to a large variability in both what and how much is taught. This therefore required analysing

the student-level impact of the curricular reform, and the influence being taught specific pedagogical content by teachers (which we refer to as adoption). While the initial focus was on student learning (see study 1 in "Study 1: student learning and the link with what teachers from the CS-schools implemented"), a parallel pilot study in grade 9 (ages 13–14) in Spring 2021 indicated that there were already significant perception-related gender gaps (El-Hamamsy et al., 2023c). This lead to the introduction of a student perception survey in Fall 2021 (see studies 2 and 3 in "Study 2: student perception, the link with what teachers from the CS-schools implemented, and correlations with performance" and "Study 3: student perception between CS-schools and schools where teachers were not yet trained to teach computer science") to determine when gender gaps appear and whether teaching CS contributes to closing these gaps.

Study 1: student learning and the link with what teachers from the CS-schools implemented
Methodology (study 1)
Participants and data collection (study 1)
The first study follows all the grade 3–4 students from 7 CS-schools over 6 months to evaluate learning in a pre- post-test design. These students were all introduced to CS for the first time during the 2018–2019 academic year and therefore had approximately 2 years of prior CS experience. The objective of the study was therefore to see to what extent these students progressed over that time period in relation to what they were taught. Given the scale of the study, the objective was to focus on a subset of the learning objectives that could be measured in a valid and reliable way, and at a large scale, in grades 3–6. We therefore chose to focus on the CT-concepts defined

---

[2] The up-to-date Computer Science curriculum can be accessed at https://www.plandetudes.ch/web/guest/education-numerique

[3] The 2021–2022 version of the pedagogical content can be accessed at https://www.vd.ch/fileadmin/user_upload/accueil/Communique_presse/decodage.pdf.
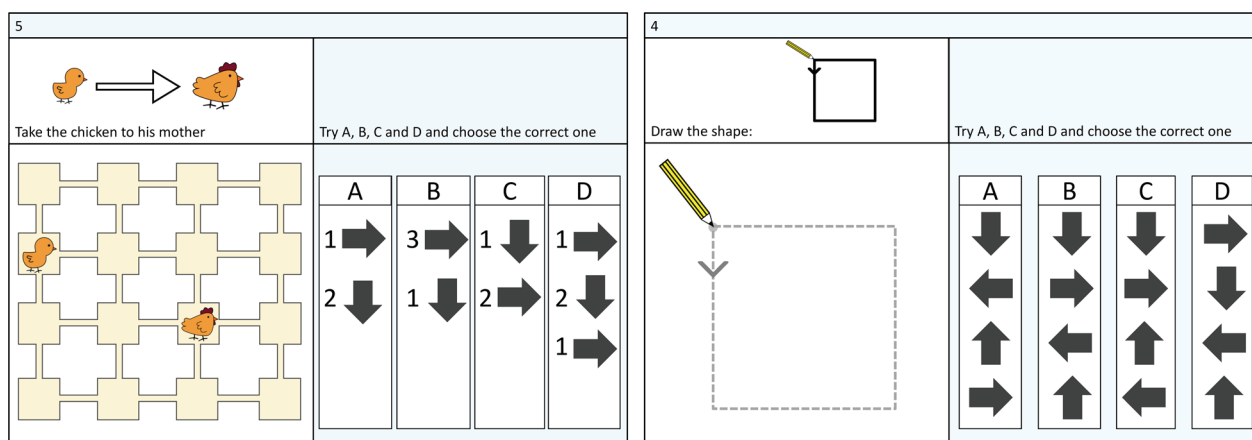
El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 6 of 43

Fig. 1 Two cCTt question formats: grid on the left and canvas on the right (figure taken from El-Hamamsy et al. 2022c)

by Brennan and Resnick (2012)[4] which align with the region's CS curricular objectives (sequences, loops, conditionals, and while statements), all the while considering what the teachers taught between the pre- and post-tests. To that effect, we employed the competent Computational Thinking test (cCTt, El-Hamamsy et al., 2022c), a 25-item CT-concepts' assessment (see example questions in Fig. 1) originally developed and validated for grades 3–4 that evaluates CS concepts of sequences, loops, if-else statements and while statements. This instrument was later validated for grades 3–6, including a Differential Item Functioning analysis which demonstrates that the cCTt is not biased towards genders (i.e. it is gender fair) and can therefore be used to measure significant differences between boys' and girls' responses (El-Hamamsy et al., 2023d).

The student-learning data were complemented by data on teachers' perception of CS and the CS-PD acquired in January 2021, and data regarding what teachers taught (which we refer to as adoption) between January and June 2021 (see Table 2). The adoption data are based on the activities that the teachers were introduced to during their CS professional development programme and is collected in the form of a number of periods per activity which we are then able to convert into boolean values and derive the amount of CS activities taught.

Please note that the data sets include missing data due to (i) students not being present for either the full pre- and/or post- tests, (ii) teachers not administering the test, or (iii) teachers not answering the pre- and/or post-teacher survey. As the analyses combine multiple data sets, a synthesis of the number of students and teachers for which the full responses are available with respect to the data subsets considered is provided in Table 3. Finally, while it would have been interesting to have a control group to be able to infer how learning compared between students who had access to CS courses and those who did not, the administration of a performance assessment to students in non-CS-schools was not authorised due to ethical concerns. Nonetheless, given the variability in what the teachers taught, 4 grade 3 classes and 6 grade 4 classes did not receive any CS education and thus provide an interesting point of comparison. As the second data subset (test + adoption data) constitutes the core of the analysis, we provide more detailed demographics information in Appendix A.1 in Table 10.

Analysis methodology (study 1)

The student learning data are analysed in three stages.

First, the January and June test data ($n = 1319$) are analysed using multiple ANOVA with Benjamini–Hochberg p-value correction to reduce the false discovery rate (study 1a). The results are reported as significant (i.e. $p < 0.05$) only if the minimum effect size (Cohen's D[5]) required to achieve a statistical power of 0.8 is reached with $\alpha = 0.05$. Dunn's post hoc test is then applied for multiple comparisons when significant. When comparing responses between groups of students (according to the dependent variables) the delta between the average

---

[4] Brennan and Resnick (2012)'s operational definition of CT decomposes CT into CT concepts (i.e. the concepts that computer scientists engage with), practices (i.e. the processes they employ to resolve computational problems) and perspectives (i.e. their perception of CT). Please note that at the time of the study there were no valid, reliable and scalable instruments to measure CT-practices and perspectives.

[5] Cohen's D is a means of quantifying the difference between the means of two samples (μ_1, μ_2) all the while accounting for their standard deviations (sd_1 and sd_2). Cohen's D is therefore computed as the difference between the two sample's means divided by the pooled standard deviation (s_p). Therefore, Cohen's $D = \frac{\mu_1 - \mu_2}{s_p}$ where $s_p = \sqrt{\frac{sd_1^2 + sd_2^2}{2}}$. The rule of thumb to interpret Cohen's D is as follows: if around 0.2 the effect is considered small, if around 0.5 the effect is considered medium and if around 0.8 or above the effect is considered large.

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 7 of 43

Table 2 Teacher survey questions (7-Point Likert, excepted adoption and demographic questions)

Dimension	References	Concept	Original question (when different from the final question)	Adaptation to the context of the curricular reform (translated from French)	M±SD ([min, max])
Professional development programme perception (Cronbach's α = 0.88)				Root: overall, the CS PD programme	
	El-Hamamsy et al. (2021b)	Interest		Was rich and interesting	1.58 ± 1.11 ([−2.0, 3.0])
	El-Hamamsy et al. (2021b)	Difficulty		Had an adapted difficulty	1.2 ± 1.27 ([−2.0, 3.0])
	El-Hamamsy et al. (2021b)	Equilibrium		Had a good equilibrium between theory and practice	1.48 ± 1.16 ([−1.0, 3.0])
	El-Hamamsy et al. (2021b)	Workshops	Not applicable	I enjoyed the workshops	2.05 ± 1.06 ([−2.0, 3.0])
	El-Hamamsy et al. (2021b)	Trainers	Not applicable	I appreciated the trainers	2.26 ± 0.79 ([0.0, 3.0])
	El-Hamamsy et al. (2021b)	Exchanges	Not applicable	I enjoyed the exchanges with the other participants	1.8 ± 1.07 ([−1.0, 3.0])
	Danaher and Haddrell (1996)	PD recommendation	Not applicable	I would recommend the CS PD programme to other teachers	1.41 ± 1.29 ([−3.0, 3.0])
	Danaher and Haddrell (1996)	Pedagogical activity recommendation	Not applicable	I would recommend the pedagogical activities seen in CS training to other teachers	1.7 ± 1.18 ([−2.5, 3.0])
CS perception (Cronbach's α = 0.93)				Root: Today I believe that CS	
	Roche (2019)	CS school mission		Is part of the school's mission	0.66 ± 1.41 ([−3.0, 3.0])
	Roche (2019)	CS transversal utility		Is useful for learning in other subjects	0.92 ± 1.22 ([−2.5, 3.0])
		CS global-utility		Is generally useful	0.94 ± 1.14 ([−2.5, 3.0])
	Deci et al. (1989)	CS potential for advancement	Not applicable	Enables students to have a better chance of professional integration	0.4 ± 1.23 ([−2.0, 3.0])
	Deci et al. (1989)	CS potential for advancement	Not applicable	Enable students to become digital actors rather than consumers	0.69 ± 1.2 ([−2.0, 3.0])
	Deci et al. (1989)	CS potential for advancement	Not applicable	Allows students to express their creativity	−0.07 ± 1.42 ([−3.0, 3.0])
		CS global-non-utility		Is not really worth it	−1.25 ± 1.24 ([−3.0, 2.0])
		CS non-utility		Will not necessarily bring much to the students	−0.86 ± 1.24 ([−3.0, 2.0])
Autonomous motivation to teach CS (Cronbach's α_all = 0.68, α = 0.76 without Introjected regulation 1) based on the Situational Motivational Angot (2013) and Motivation at work Gagné et al. (2010) scales				Root: The reason I have implemented/plan to implement CS activities with my students is that	

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 8 of 43

Table 2 (continued)

Dimension	References	Concept	Original question (when different from the final question)	Adaptation to the context of the curricular reform (translated from French)	M±SD ([min, max])
	Angot (2013)	Intrinsic motivation 1	I find this activity really pleasant.	Teaching CS is fun	0.65 ± 1.28 ([−3.0, 3.0])
	Angot (2013)	Identified regulation 1	I think this activity is important to me.	I believe that teaching CS is important for students	0.84 ± 1.26 ([−2.0, 3.0])
	Angot (2013)	External regulation 1	I feel I am supposed to do it.	I feel I am supposed to do this	0.75 ± 1.61 ([−3.0, 3.0])
	Angot (2013)	Intrinsic motivation 2	I find this activity interesting	I find that teaching CS interesting	0.87 ± 1.18 ([−3.0, 3.0])
	Angot (2013)	Identified regulation 2	I find that doing this activity is good for me.	I find that teaching CS is useful for my students	0.9 ± 1.2 ([−2.5, 3.0])
	Angot (2013)	External regulation 2	It's something I have to do.	I feel that this is something I have to do	0.7 ± 1.37 ([−3.0, 3.0])
	Gagné et al. (2010)	Introjected regulation 1	My work is my life and I don't want to fail	I want to show others that I can do it	−1.35 ± 1.62 ([−3.0, 2.0])
	Gagné et al. (2010)	Introjected regulation 2	My reputation depends on it	My reputation depends on it	−1.54 ± 1.54 ([−3.0, 3.0])
Adoption	El-Hamamsy et al. (2022a)	Number of activities		Which activities did you teach? [List of all the CS pedagogical activities the teachers were trained to introduce in the CS-PD programme they participated in in 2018–2019 for which they had access to all the requires material and pedagogical resources]	2.3 ± 1.88 ([0.0, 7.0])
	El-Hamamsy et al. (2022a)	Periods		How many periods did you teach per activity? [List of the activities that the teachers selected in the previous question]	9.15 ± 10.36 ([0.0, 45.0])
Demographics		Age		Age	39.44 ± 11.51 ([23.0, 62.0])
		Teaching experience		Years of teaching experience	15.24 ± 10.91 ([0.5, 38.0])
		ICT experience		Years of experience with informatics	11.78 ± 8.15 ([0.0, 30.0])
		Teaching Digital Education experience		Years of experience in teaching digital education	2.49 ± 2.22 ([0.0, 18.0])
		Perceived ICT competence		When it comes to ICT, I consider myself to be (1 = a non-user, 2 = a novice, 3 = an intermediate, 4 = somebody who is used to it, 5 = an expert)	3.32 ± 0.75 ([1.0, 5.0])

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 9 of 43

Table 2 (continued)

Dimension	References	Concept	Original question (when different from the final question)	Adaptation to the context of the curricular reform (translated from French)	M±SD ([min, max])
		Relative ICT competence		I tend to use digital technologies (1 = after most of my colleagues, 2 = as the same time as most of my colleagues, 3 = before most of my colleagues, I am an early adopter, 4 = before anybody else, I am an innovator)	2.17 ± 0.69 ([1.0, 4.0])

Cronbach's α is provided for each sub-scale. Please note that the sample size ($n = 67$) was too small to validate the measurement model through Confirmatory Factor Analysis

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 10 of 43

Table 3 Number of students participating in study 1 on student learning structured according to the number of complete observations according for each data subset considered: pre- (January) and post-test (June) data, teacher adoption data (at the time of the post-test, June), teacher perception data (at the time of the pre-test, January)

Data subset	Grade	Number of students per grade		
		Boys	Girls	Total
0. Pre-test (January) or post-test (June) data	Grade 3	357	313	670
($n_{classes} = 83$)	Grade 4	363	351	714
	All grades	720	664	1384
1. Pre-test (January) or post-test (June) data	Grade 3	332	297	629
($n_{classes} = 74$)—study 1a	Grade 4	353	337	690
	All grades	685	634	1319
2. Pre-test (January) or post-test (June) data and adoption data	Grade 3	256	224	480
($n_{classes} = 55$)—study 1b	Grade 4	265	244	509
	All grades	521	468	989
3. Pre-test (January) or post-test (June) data and adoption data and perception data	Grade 3	207	186	393
($n_{classes} = 43$)—study 1c	Grade 4	207	173	380
	All grades	414	359	773
4. Pre-test (January) or post-test (June) data and perception data	Grade 3	257	238	495
($n_{classes} = 55$)—study 1c	Grade 4	262	232	494
	All grades	519	470	989

scores on the cCTt's scale is provided (Δ), in addition to the F-value, degrees of freedom, corresponding p-value and effect size using Cohen's D. The ANOVA considers the students' scores as the dependent variable, and the interaction between the following independent variables: time (pre-test or post-test), grade (3 or 4) and gender (boy or girl as indicated on the school's records[6]).

Second, the data set that introduces the adoption data ($n = 989$), i.e. what the teachers taught between the pre- and post-test, is analysed through hierarchical linear modelling which nests students in classes and classes in schools (study 1b).

Finally, to determine whether teacher-level variables (see Table 2) influence student learning, the third data set (study 1c) that includes teacher perception is analysed through a correlation analysis with averaged class-level student scores ($n = 67$), prior to a hierarchical linear modelling at the student-level ($n = 752$). The hierarchical linear modelling done in these two stages was conducted in R (version 4.2.1, R Core Team, 2019) with nlme (version 3.1-157, Pinheiro et al., 2022; Pinheiro & Bates, 2000) and sjstats (version 0.18.2, Lüdecke, 2022).

Results: the impact of teaching CS on student learning (study 1)

Student learning and the influence of gender and when the test was taken (study 1a)

The ANOVA indicates that all independent variables and their interactions significantly influence the test score (see Appendix A.3 Table 12 for a synthesis of the effects) and the following trends emerge.

---

[6] Please note that we never asked students to relate their gender throughout the data collection process to avoid biasing students' responses and performance as a result of stereotype threat. Indeed, as we could not guarantee that all students would participate in all the data collections which were conducted over multiple sessions, and therefore could not solely rely on collecting the gender information at the end of the final data collection, we relied on the gender information obtained from the school's records. This information is provided by students' parents to the schools and therefore most likely aligns with the students' sex, without a guarantee that this corresponds to up-to-date information regarding the way students identify themselves. Furthermore this gender information was provided by the schools in a binary format. Although we acknowledge that gender relates to a person's identity, differs from biological sex (Risman, 2018), and is increasingly recognised as being non-binary, this was not yet fully the case in the country where the study was conducted at the level of formal primary education and at the time of the data collection. Indeed, at the time of the data collections, gender at the level of primary school and formal education more broadly was mainly considered as a binary construct. Nonetheless, most international studies find that the proportion of people who identify as transgender is generally inferior to 1.5% (e.g. 0.6% of the population aged 13 or older in the US, Herman et al. (2022); between 0.5% and 1.3% for children, adolescents and adults according to Zucker 2017's international review).

Footnote 6 (continued)

The potential discrepancy between the gender information on the school's records and students' gender identity represents therefore at most a 1.5% error which is below the level of significance which would affect the validity of the findings with a confidence level $\alpha = 0.05$. Therefore, in order to align with the current practice in the STEM education community which often employ the term gender and gender biases when actually gathering and analysing binary or biological sex data (e.g. Jensen et al., 2023; Sung et al., 2023; Malespina & Singh, 2023), we maintain the term gender, gender-biases and gender-gaps when referring to our data and our analyses.
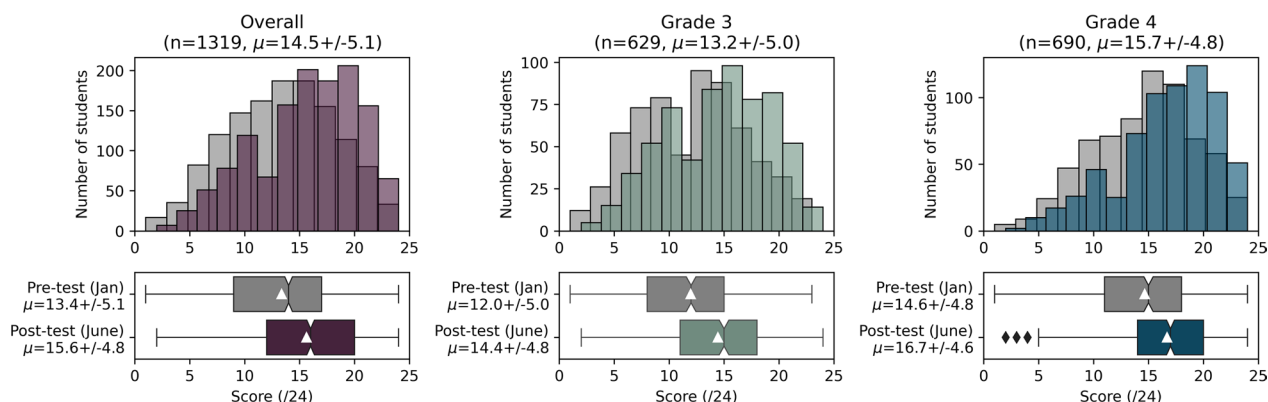
El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 11 of 43

Fig. 2 Student performance distribution according to grade and whether in the pre- or post-test

Are all students progressing?

As Fig. 2 shows, grade 4 students perform better than grade 3 students with a medium effect size overall (grade 4>3 $\Delta = 2.468$, $p < 0.0001$, Cohen's $D = 0.502$) , in the pre-test (pre-test grade 4 > 3, $\Delta = 2.686$, $p = 0.0$, Cohen's $D = 0.549$) and in the post-test (post-test grade 4 > 3 $\Delta = 2.249$, $p = 0.0$, Cohen's $D = 0.482$). Students also performed better on the post-test overall (post-test > pre-test, $\Delta = +2.256$, $p < 0.0001$, Cohen's $D = 0.457$) with students in grades 3 and grade 4 improving by a medium effect size (grade 4 post > pre, $\Delta = 2.048$, $p = 0.0$, Cohen's $D = 0.436$; grade 3 post > pre, $\Delta = 2.485$, $p = 0.0$, Cohen's $D = 0.51$). Interestingly, the grade 3 students' performance in the post-test (June) was equivalent to the grade 4 students' performance on the pre-test (January), although only 6 months separated the assessments (grade 4 pre-test $\sim$ grade 3 post-test, $\Delta = 0.201$, $p = 0.4444$, Cohen's $D = 0.042$).

Are there gender biases and are these closing?

The results that account for the students' gender alone show that there is a significant main effect of students' gender on their performance. In particular, boys have significantly higher scores than girls overall with a small effect size (boys > girls, $\Delta = 0.551pts$, $p = 0.0015$, Cohen's $D = 0.109$). Considering the two-way interaction effects, we observe the following tendencies. Over all students, the gender gap is significant in the pre-test (January boys > girls, $\Delta = 0.664pts$, $p = 0.0079$, Cohen's $D = 0.131$) but decreases and is no longer significant by the post-test (June boys $\sim$ girls, $\Delta = 0.438pts$, $p = 0.0744$, Cohen's $D = 0.091$). Considering the two-way interactions, these gender differences are significant in grade 3 (grade 3 boys > girls, $\Delta = 0.725pts$, $p = 0.004$, Cohen's $D = 0.145$) , but not in grade 4 (grade 4 boys $\sim$ girls, $\Delta = 0.469pts$, $p = 0.0604$, Cohen's $D = 0.098$) . The three-way interaction between these variables thus helps

shed some light on the trends observed (see Fig. 3) to draw conclusions:

- In grade 3 there is a small marginally significant gap in the pre-test (grade 3 pre-test boys $\sim$ girls, $\Delta = 0.764pts$, $p = 0.0526$, Cohen's $D = 0.161$) and a small significant gap in post-test (grade 3 post-test boys > girls, $\Delta = 0.687pts$, $p = 0.0422$, Cohen's $D = 0.139$) , with the effect sizes indicating that the gap is getting smaller, but has not yet closed.
- In grade 4 there are small marginally significant differences in the pre-test (grade 4 pre-test boys $\sim$ girls, $\Delta = 0.727pts$, $p = 0.0624$, Cohen's $D = 0.151$) and no significant differences observed in the post-test (grade 4 post-test boys $\sim$ girls, $\Delta = 0.211pts$, $p = 0.5046$, Cohen's $D = 0.046$) , indicating that the gender gap has closed.

To complement these findings we consider the student learning data from study 2 (see "Study 2: student perception, the link with what teachers from the CS-schools implemented, and correlations with performance") that was conducted in November 2021 (5 months after the post-test of study 1) in the same schools and includes students from grades 3–6 (7–11). This is a particularly interesting cohort of students because students in grades 3 and 4 in study 2 are the first group of students to have had access to CS education starting first grade. Analysing the student performance data confirms that students continue to progress in terms of CT-concepts when moving on to grades 5 and 6 (see Fig. 4). Indeed, the differences between grades 3 and 4 are significant ($\Delta = 2.87pts$, $p < 0.0001$, Cohen's $D = 0.566$), as well as those between grades 4 and 5 ($\Delta = 1.35pts$, $p < 0.0001$, Cohen's $D = 0.266$), although there is no significant difference between students in grades 5 and
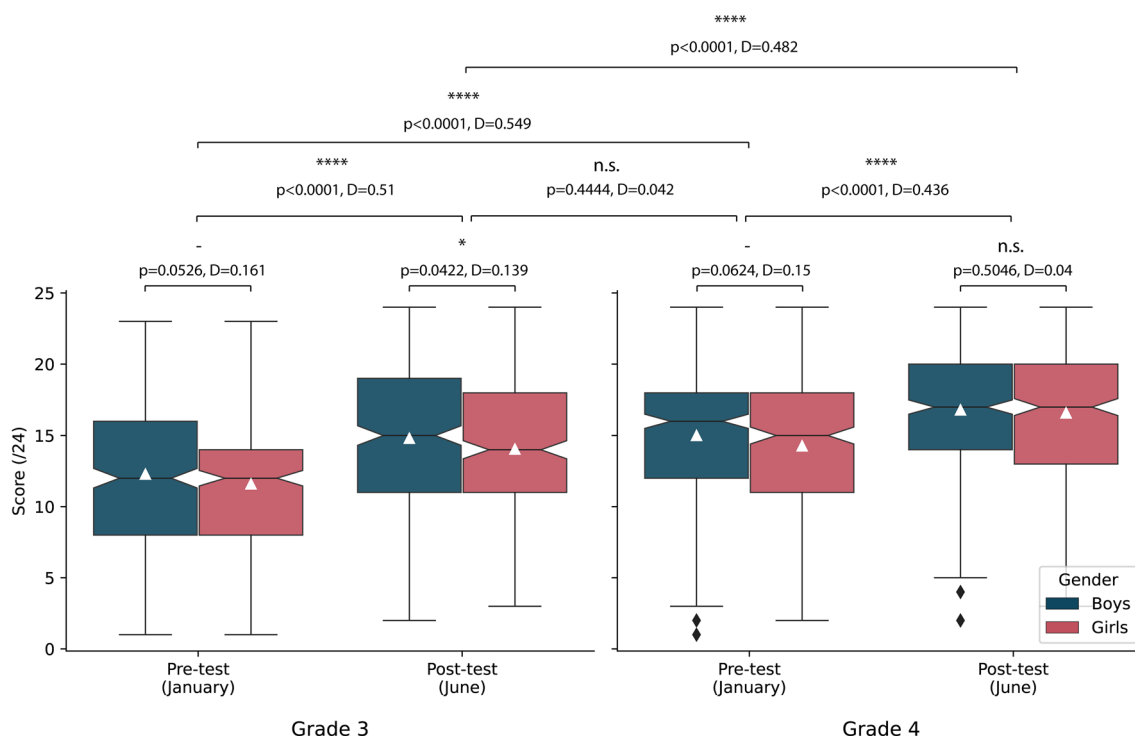
El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 12 of 43

Fig. 3 Student performance distribution according to grade, gender and whether in the pre- or post-test

6 ($\Delta = 0.423pts$, $p = 0.1345$, Cohen's $D = 0.083$). This is consistent with the fact that students increase in maturity faster when they are younger (Hartshorne & Germine, 2015). As such, students in grades 3 and 4 differ more significantly in terms of their cognitive abilities than students in grades 5-6.

Evaluating the difference between boys' and girls' scores per grade in November 2021 (study 2) indicates that the results are non-significant across grades (see Fig. 4). As these students were in their 3rd or 4th year of CS education, this would appear to corroborate the previous findings: students who have had early and prolonged access to CS education are less likely to exhibit CS-performance gender-gaps.

Student learning and the influence of the CS-education received (study 1b)

To understand how teaching the CS-pedagogical content from the curriculum may have influenced student learning, we consider the data from 989 students for whom the pre- and post- tests, and teacher adoption data (i.e. what the teachers taught, see "Participants and data collection (study 1)") are available. We implemented multiple hierarchical linear models while nesting students in classes and classes within schools to account for the

different ways of considering student learning and adoption[7] These models consistently indicated that there was no direct link of adoption on students' post-test scores. For instance the model considering how the delta between the post and pre-tests is influenced by the students' grade, gender and the number of CS activities taught estimates a non-significant effect of the number of CS activities taught on the progress students made with $b = 0.122$, df $= 45$, $t = 0.442$, and $p = 0.661$ (see Table 11 in Appendix A.2). Only the pre-test score significantly predicts the progress made in the post-test, with students performing lower at the pre-test progressing more. While the lack of a significant influence of CS activities taught on learning may appear surprising, visualising the trends between teaching and not teaching CS pedagogical content, as well

---

[7] The hierarchical linear models considered the following:
- Dependent variables: the delta between the post-test and pre-test scores or the normalised change (a symmetrical version of the learning gain, Coletta and Steinert, 2020)
- Independent variables: the interaction between pre-test score, grade (3 or 4), and different adoption metrics (number of activities, or amount of CS-education time)
- Random effects: classes within schools. Please note that random effects are not the main focus of the analysis but still need to be included in the hierarchical linear model in order to account for their influence on the dependent variables. We therefore do not estimate the impact of each school or class on the outcome but rather control for them in order to avoid drawing erroneous conclusions.
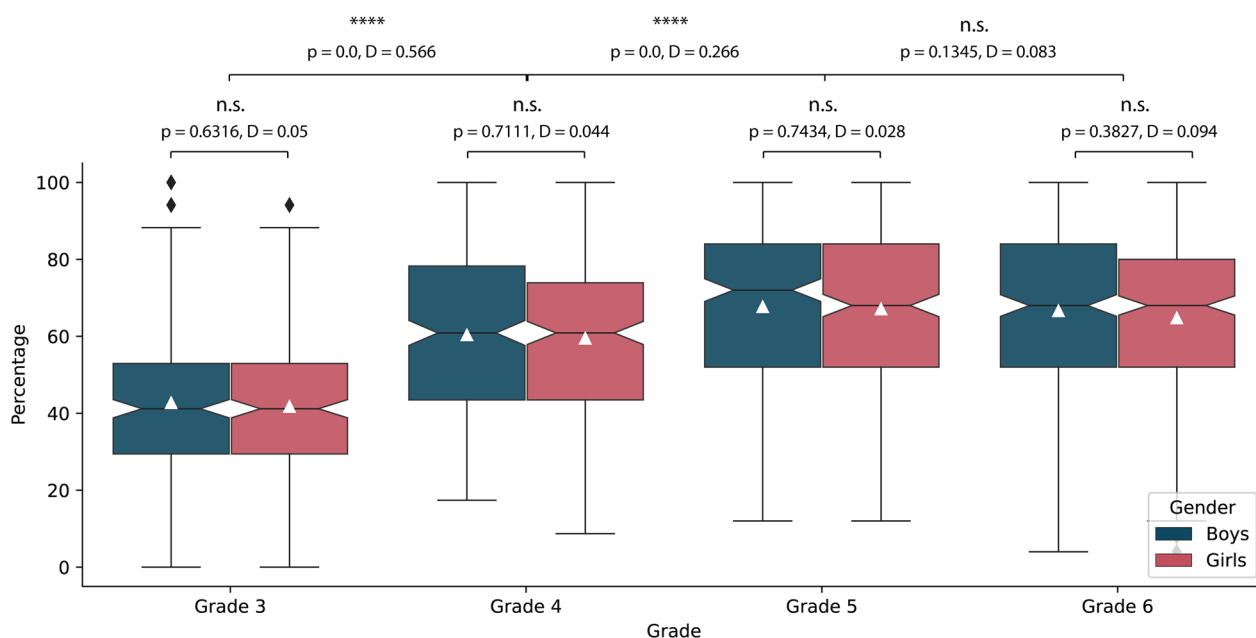
El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 13 of 43

Fig. 4 Student performance distribution according to grade and gender using data from the second study (*n* = 2226, November 2021). All grade-differences are significant, excepted the one between grades 5 and 6 while the gender-differences per grade are non-significant

as according to the number of activities taught, confirms the lack of an evident trend (see Fig. 5).

Student learning and the influence of teacher demographics, perception and the CS-PD received (study 1c)

Given the link between access to CS education and performance, and the lack of a direct link between what the teachers taught and student learning, it would appear that there are additional factors at play when affecting learning. Therefore in a final phase, the teachers' aggregate (i) perception of the PD programme, (ii) perception of CS, (iii) autonomous motivation to teach CS[8] and the (iv) demographic data collected at the same time as the pre-test were put in relation to the student learning results. First, the students' results were averaged per class to obtain a class performance and correlated with the teacher-level variables. As the perception data are on a 7-point Likert scale and non-normally distributed, Spearman's rank correlation was used. All the correlations with class performance were non-significant (whether in terms of teacher demographics, prior experience or CS

perception), with the exception of the training evaluation (Spearman's rho = 0.33, *p* = 0.007).

As adoption was found to be not significantly related to student learning (study 1b), we compared two hierarchical linear models at the student-level, one with and one without adoption variables, with both including student-level, teacher perception-level and teacher demographic-level variables. An analysis of variance between the two models indicates that the difference is non-significant (*p* = 0.768). The more parsimonious model which does not include the adoption data (see Table 4), and which also relies on a larger set of complete data (i.e. 1027 vs. 752 observations) should therefore be preferred. The resulting hierarchical linear model at the student-level confirms the trend observed in the correlation analysis, and indicates that the following dependent variables predict the delta between the pre- and post-test scores, with no influence of teacher demographic variables (including teaching and ICT experience):

- The pre-test score predicts the delta negatively (*p* < 0.0001, $\beta = -0.35$), i.e. students performing lower at the pre-test progressed more.
- The average PD programme evaluation score predicts the delta positively (*p* = 0.0053, $\beta = 1.02$), i.e. students of teachers who positively viewed the CS-PD progressed more.

---

[8] The Autonomous Motivation (AM) score is computed using the Relative Autonomy Index (Grolnick & Ryan, 1989) by combining the sub-scales for intrinsic motivation (IM), identified regulation (IdR), introjected regulation (InR) and external regulation (ER) and aggregating them as explained by Howard et al. (2020). That is to say: $AM = (2 \times IM + 1 \times IdR - 1 \times InR - 2 \times ER)/6$
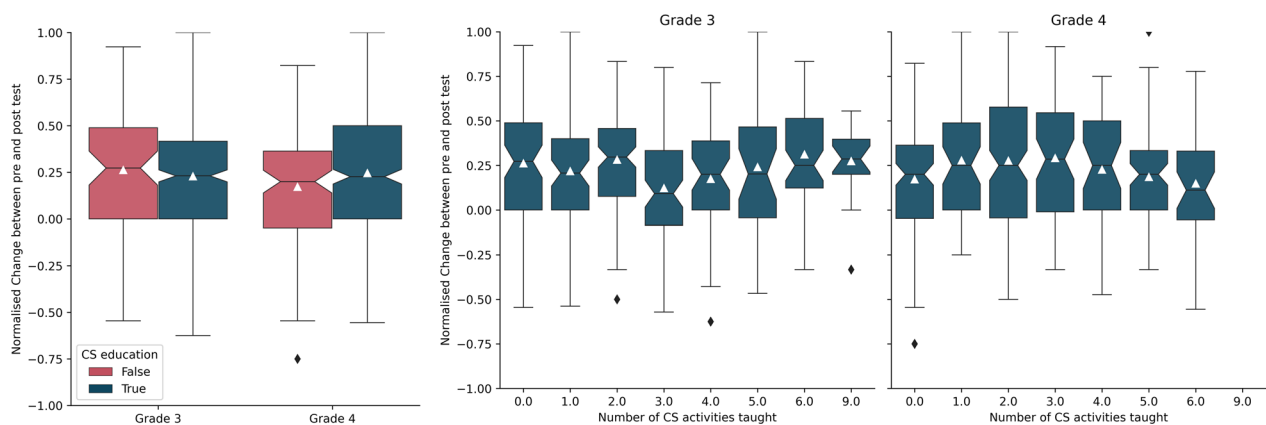
El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 14 of 43

Fig. 5 Student normalised change distribution according to grade, access to CS-education (left) and the number of CS-activities taught (centre for grade 3, right for grade 4). A two-way ANOVA between the grade and what was taught does not identify any significant differences between groups in terms of access to CS education ($F(2) = 1.05$, $p = 0.35$). A one-way ANOVA per grade did not identify any significant differences according to the number of activities taught (grade 3 $F_3(1) = 0.13$, $p_3 = 0.72$; grade 4 $F_4(1) = 0.89$, $p_4 = 0.35$)

Table 4 Hierarchical linear model for student learning with respect to student-, and teacher-level variables (dependent variable: delta between pre- and post-test scores, $n = 1027$ students in 57 classes in 6 schools)

Variables		Estimate	95% CI	Std. error	Degrees of freedom	*t*-value	*p*-value
Student-level	**(Intercept)**	**10.48**	[5.42, 15.54]	2.579	968	4.06	**0.0001**
	Pre-test score	**−0.35**	[−0.40, −0.31]	0.023	968	−15.54	**0.0000**
	Gender (girl)	0.17	[−0.22, 0.55]	0.198	968	0.84	0.4021
	Grade (3)	0.35	[−0.34, 0.53]	0.823	39	0.42	0.6746
	Grade (4)	1.15	[−0.13, 0.73]	0.814	39	1.41	0.1663
Teacher-perception	**CS-PD programme evaluation**	**1.02**	[0.07, 0.38]	0.344	39	2.96	**0.0053**
	CS utility perception	0.15	[−0.84, 1.06]	0.628	39	0.24	0.8122
	CS non-utility perception	0.27	[−0.72, 1.18]	0.600	39	0.45	0.6561
	CS autonomous motivation	−0.21	[−0.97, 0.56]	0.380	39	−0.54	0.5896
Teacher demographics	Age	−0.14	[−0.30, 0.01]	0.078	39	−1.83	0.0743
	Experience with informatics	0.01	[−0.06, 0.09]	0.036	39	0.41	0.6828
	Teaching experience	0.11	[−0.05, 0.27]	0.079	39	1.42	0.1640
	Digital education teaching experience	−0.10	[−0.31, 0.11]	0.104	39	−0.94	0.3552
	Perceived ICT competence	−0.57	[−1.34, 0.19]	0.376	39	−1.53	0.1345
	Perceived relative ICT competence	0.20	[−0.65, 1.05]	0.421	39	0.47	0.6387

Significant variables are highlighted in bold. $R^2 = 0.279$, $AIC = 5386$, $BIC = 5474$, $RMSE = 3.04$. Random effects $\sigma^2 = 9.72$, $\tau_{class} = 0.57$, $\tau_{school} = 1.37$

Please note that (i) the classes had an average of 18 ± 2 students per class (minimum 14, maximum 22); (ii) the schools had an average of 8 ± 5 classes (i.e. 8, 8, 1, 17, 6, 7, 8 classes) who participated in the three data collections required for this analysis. These numbers are coherent with the relative sizes of the schools, with the exception of the third where the majority of teachers chose not to participate in the data collection

Synthesis and limitations of study 1

The students progress in terms of CT-concepts over time, with grade 3 students achieving a year's worth of CT-development in a 6-month window (study 1a, *positive impact*). However, the results of the hierarchical linear modelling indicate that there is no direct effect of what was taught with the progress students made (study 1b, *no impact and therefore negative for equity*). The only factors that appear to influence learning are: (i) the students' scores in the pre-test, with students who have lower scores progressing more thus contributing to *performance-equity*; (ii) the teachers' perception of the PD programme (study 1c, *positive impact*). There is additionally no influence of teachers' demographics on what the students have learnt, indicating that the PD programme helped prepare teachers to teach CS pedagogical content, irrespective of their prior teaching experience and ICT experience. This contributes once

more to *equity* by ensuring that all students have access to quality CS education, irrespective of the teachers' background (structural barriers). Finally, the findings indicate the existence of gender gaps (study 1a, likely due to social barriers) but that these get smaller the longer students are in contact with CS education (positive for *gender-equity*).

There are, however, limitations due to the lack of a true control group that has never had access to CS education. Indeed, the students in the present study were not compared to students who had not done any CS education between the pre- and post-tests, or since the start of their schooling. The fact that students with lower pre-test scores progress more may also be due to the existence of a "ceiling effect" for already higher performing students (either cognitively, with respect to what the cCTt measures, or what is attainable with the pedagogical content taught). In terms of teacher and class data, while the teachers were asked what they taught and for how long, this does not indicate their mastery of the content, the implementation fidelity (i.e. to what extent they put emphasis on the CS concepts in these activities) or whether they taught other activities that were not part of the PD programme that may be linked to CS education or grid based concepts which are also part of the maths curriculum. Finally, the assessment used:

- focuses on CT-concepts, although there are other elements of CT that may be positively affected by access to CS education which are not measured (in addition to other dimensions of the CS curricular reform including those pertaining to machines and networks, data and information and the impact of CS on society);
- is used in both the pre- and post-test due to the fact that (i) at the time of the studies there existed no valid and reliable assessment of CT-concepts in primary school for these grades; (ii) no validated assessment proposes isomorphic variants which have been proven to have the exact same difficulty and can therefore be reliably employed in the comparison of pre–post test design. To the best of our knowledge this remains true today as only Parker et al. (2022) has begun investigating how to create an isomorphic version of their instrument (the ACES test) and analysed what types of changes to the questions could truly be considered isomorphic in this context. This is important because "seemingly superficial changes in an item's context can cause students to recruit different knowledge and cognitive processes when solving a problem" (Parker et al., 2022).

Study 2: student perception, the link with what teachers from the CS-schools implemented, and correlations with performance
Methodology (study 2)
Participants and data collection (study 2)
This study extends the first by evaluating students' mastery of CT-concepts and their perception of the discipline. The data collection was conducted in November 2021 and involved all students from grades 3–6 in the 7 CS-schools involved in the first study (see Table 5). The students first responded to a perception survey, before being administered the cCTt (which was shown to be adapted for grades 5–6 in El-Hamamsy et al. 2023d) to assess their mastery of CT-concepts.

The perception survey (see Table 6) targeted three dimensions.

The first dimension is the students' perception of *Computer Science*, including who they perceive as doing CS, called "informatics" in the region, a scalable alternative to the draw-a-computer-scientist test (Pantic et al., 2018). Students were asked whether they perceived certain role models (e.g. influencers such as parents and teachers, Wang & Hejazi Moghadam, 2017), someone else, or nobody, as doing CS. One hypothesis is that students who have access to CS-education are more likely to perceive their teachers as role models. As primary school teachers are mainly women, they can be considered female role models, an element that is key to engaging girls in the field (Cheryan et al., 2017; Kong et al., 2018). Another hypothesis is that perceiving people "close to them" as doing CS (i.e. related to the idea that CS is becoming ubiquitous and accessible to all), will contribute to improved perception of CS overall.

The second dimension is how students perceive *robots*, as robotics is a means of teaching CS (El-Hamamsy et al., 2021a), and CS and engineering tend to be subject to stronger stereotypes than science and maths among young students (Master et al., 2017). Interestingly, recent studies have found that there is a link between students' perception of robots and their "aspirations to pursue a career in science" (Giang et al., 2023), with introductions to educational robotics affecting their perception of robots.

The third dimension is how students perceive *tablets and other digital devices* which are also employed as means of teaching CS (and ICT) in the curricular reform.

For each of these dimensions (CS, robotics, tablets), the emphasis is placed on three factors that are "different but related aspects of motivation" (Master et al., 2017) and can be considered as predictors of academic achievement in general (Bandura, 1993; Howard et al., 2021a; Olivier et al., 2019), educational choices, and career decisions (Blotnicky et al., 2018; Mason & Rich, 2020; Wang et al.,

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 16 of 43

Table 5 Number of students participating in the first perception survey and the third test (study 2, November 2021) and their intersection with the teacher adoption survey

Subset	Gender	Grade				Total
		3	4	5	6	
Perception ($n_{classes} = 142$)	Boys	263	307	314	334	1218
	Girls	265	286	311	328	1190
	All	528	593	625	662	2408
Perception and adoption ($n_{classes} = 114$)	Boys	196	230	285	351	1062
	Girls	201	220	287	346	1054
	All	397	450	572	697	2116
Test ($n_{classes} = 140$)	Boys	240	282	289	317	1128
	Girls	243	252	296	307	1098
	All	483	534	585	624	2226
Test or perception ($n_{classes} = 142$)	Boys	265	311	317	337	1230
	Girls	272	287	314	330	1203
	All	537	598	631	667	2433
Test and perception ($n_{classes} = 139$)	Boys	209	262	270	295	1036
	Girls	214	239	281	294	1028
	All	423	501	551	589	2064
Test and perception and adoption ($n_{classes} = 105$)	Boys	166	198	285	267	916
	Girls	175	182	287	252	896
	All	341	380	572	519	1812

2020), in addition to being the most prominent in surveys evaluating students' (at all levels of education) perception of CS, coding or STEM (Mason & Rich, 2020):

- Interest, i.e. "how much the individual likes or is interested in the activity" (Mason & Rich, 2020), is a key component of intrinsic motivation in self-determination theory (Ryan & Deci, 2020) and expectancy-value theory (Eccles & Wigfield, 2020). Several studies have found that boys are more interested in CS than girls, as in most STEM-related disciplines (Mason & Rich, 2020). Comfortingly, researchers have also found that interest increases after access to CS experiences, in particular for girls, which contributes to closing the interest gender gap (Master et al., 2017).
- Self-efficacy (Bandura, 1993; Kong et al., 2018), i.e. "a person's belief that they can complete a particular task or fulfil a particular role within a specific domain" (Mason & Rich, 2020). Similarly to interest, self-efficacy has been found to be higher for boys than girls in STEM-related domains. Self-efficacy has also been found to increase with computing experience, in some cases even contributing to closing the gender gap (Mason & Rich, 2020), whether related to programming (Gunbatar & Karalar, 2018), or robotics (Master et al., 2017).

Please note that, as domain-specific self-efficacy may be related to general self-efficacy, we also consider a school-related self-efficacy variable in the survey (i.e. how well students believe they are able to perform in school in general).

- Perceived utility (Eccles & Wigfield, 2020; Wigfield & Eccles, 2000), a component of expectancy-value theory referring "to how a task fits into an individual's future plans" which is considered to "directly [influence] a person's achievement-related choices, and is influenced by a person's experiences, perceptions, goals, and self-schemata" (Mason & Rich, 2020; Wigfield & Eccles, 2000).

Given that the same survey was administered from grades 3–6 in conjunction with the test, the survey needed to be short to account for the students' age and attention span (see Table 6). Cronbach's α measurement of internal consistency of scales is provided for all Likert-type questions employing an analogue visual scale (see Fig. 6). This is complemented by a Confirmatory Factor Analysis to confirm the adequacy of the complete measurement model (see "Results: perception, the influence of what was taught since the start of the year on perception, and the link with performance (study 2)"). Finally, the student survey was accompanied by a teacher adoption-survey that asked each teacher the amount of time spent

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 17 of 43

Table 6 Student perception survey items translated from French

Dimension	Concept	Question	Format
Computer Science	Interest	I like informatics	5PT-AVS
	Self-efficacy	I am capable of learning informatics	5PT-AVS
	Utility	We can do a lot of things with informatics	5PT-AVS
	Role models	When I think of someone who does informatics I think of (you can chose multiple answers):	Checkboxes
		The teacher	
		My mother	
		My father	
		A sibling or friend	
		Somebody else	
		Nobody	
Robotics	Interest	I like robots	5PT-AVS
	Self-efficacy	I am capable of using robots	5PT-AVS
	Utility	We can do a lot of things with robots	5PT-AVS
	Usage	When I am at school or at home I use or play with the following robots (you can chose multiple answers):	Checkboxes
		Thymio	
		Bluebot	
		Lego Robots (WeDo, Spike, Prime, Mindstorm or Technic)	
		Cubetto	
		mBot	
		Ozobot	
		Other robots	
Tablets	Interest	I like tablets	5PT-AVS
	Self-efficacy	I am capable of using tablets	5PT-AVS
	Utility	We can do a lot of things with tablets	5PT-AVS
	Usage	When I am at school or at home I use a tablet or computer to (you can chose multiple answers)	Checkboxes
		Take photos and videos	
		Call, text, watch videos or listen to music	
		Play games	
		Read	
		Programme (e.g. Scratch)	
		Draw, create interactive albums or music	
		Nothing	
General	School-related self-efficacy	I am capable of doing well at school	5PT-AVS

Cronbach's $\alpha_{CS, Robotcs, Tablets} = 0.67$ for the 9 items consisting of 3 sub-scales using the 5-point Analogue Visual Scale (5PT-AVS) and is considered to be between acceptable and good (George & Mallery, 2003), and to have between moderate and high reliability (Hinton, 2004)

Please note that the items pertaining to the usage of robots and tablets are not investigated in the present article

teaching each of the CS, ICT and robotics activities proposed in the PD-programm.

Please note that the survey was initially intended as a pre–post administration to be put in relation with what the students did in between (as in study 1, see "Study 1: student learning and the link with what teachers from the CS-schools implemented"). However, the positively skewed results (see "Results: perception, the influence of what was taught since the start of the year on perception, and the link with performance (study 2)") indicated

that the students' perception of the discipline was possibly impacted by the CS-education received in prior years. It was thus essential to compare with students who had not yet received any CS-education. Unlike administering an assessment of CT-concepts to students who had not received any CS-education, administering a perception survey to a control group was accepted from the ethical standpoint (see study 3 in "Study 3: student perception between CS-schools and schools where teachers were not yet trained to teach computer science").

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60
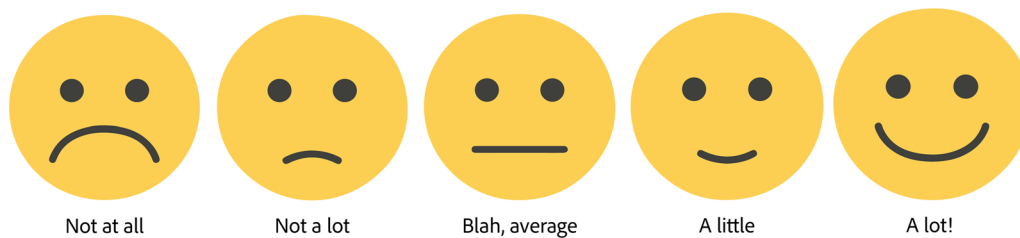
Page 18 of 43

Fig. 6 Analogue Visual Scale employed for the student survey's Likert questions. The labels in French (original survey language) were established with teachers and validated in a pilot run with two classrooms

Analysis methodology (study 2)

The analysis is conducted in three stages:

1. Descriptive analysis of students' perception of the discipline.
2. Structural Equation Modelling (SEM) to assess the impact of student demographic variables (gender, grade, general school-related self-efficacy), class-level variables (with respect to CS-, robotics- and ICT-related education received since the start of the year) on students' perception of the discipline (see Fig. 7).
3. Introducing student performance variables into the previous SEM to see how perception of the discipline may influence performance (see Fig. 8).

To assess the models' goodness of fit, both the measurement model (CFA) and the structural models (SEM) must be validated. Hu & Bentler (1999) recommend employing multiple complementary fit indices. Therefore, we employed local and global fit indices, namely the ratio between the χ^2 statistic and the degrees of freedom, the comparative fit index (CFI), the Tucker–Lewis index (TLI), the root mean square error of approximation (RMSEA) and the standardised root mean square residual (SRMR). While the χ^2 statistic should be non-significant (Alavi et al., 2020; Prudon, 2015), this is rarely the case, which is why researchers have recommended employing the ratio between the χ^2 statistic and the degrees of freedom (df). The ratio χ^2/df should be inferior to 5 for acceptable fit, and inferior to 3 for good fit (Kyriazos, 2018). The CFI and TLI should be above 0.9 for acceptable fit and above 0.95 for good fit (Byrne, 1994; Schumacker & Lomax, 2004; Xia & Yang, 2019). The RMSEA on the other hand should be below .08 for acceptable fit and below .06 for good fit (Chen et al., 2008; Hu & Bentler, 1999; Xia & Yang, 2019). Finally, the SRMR should be below 0.08 (Hu & Bentler, 1999; Xia & Yang, 2019).

As the data are not normally distributed and is positively skewed, in addition to including binary variables, the CFA and SEM analyses were conducted using robust diagonally weighted least square estimators. The modelling was conducted in R (version 4.2.1, R Core Team, 2019) with lavaan (version 0.6-11, Rosseel, Rosseel (2012)), semTools (version 0.5-6, Jorgensen et al., 2022), semTable (version 1.8, Johnson & Kite, 2020), psych (version 2.2.5, Revelle, 2022), and semPlot (version 1.1.5, Epskamp, 2022).

Results: perception, the influence of what was taught since the start of the year on perception, and the link with performance (study 2)

Students' perception of CS, robots and tablets is highly positive and nearly saturates ($M = 1.55 \pm 0.84$ on the -2 to $+2$ scale, see Fig 9). An ANOVA, however, indicates that there are small significant gender differences.

As Fig. 10 shows, boys:

- are more interested in CS ($p < 0.0001$, Cohen's $D = 0.253$),
- are more interested in tablets ($p = 0.006$, Cohen's $D = 0.117$),
- have higher tablet self-efficacy ($p = 0.0042$, Cohen's $D = 0.124$),
- perceive robots more favourably on all criteria ($p < 0.0001$, Cohen's $D = [0.197, 0.363]$).

Gender biases are also found in terms of who the students perceive as doing CS ($\chi^2(5) = 15.7$, $p = 0.008$, see Fig. 11). In particular, boys consider that their father does CS more often than girls ($\chi^2(1) = 10$, $p = 0.0017$), while girls perceive that their teacher does CS more often than boys ($\chi^2(1) = 16$, $p = 0.0001$).

To gain better insight into how the student-factors interact (demographic variables, perception of CS, tablets and robots, CS role models), and are influenced by what teachers taught, we employed SEM ($n = 2116$, November 2021) using Robust Diagonally Weighted Least Squares estimator (WLSMVS).

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 19 of 43
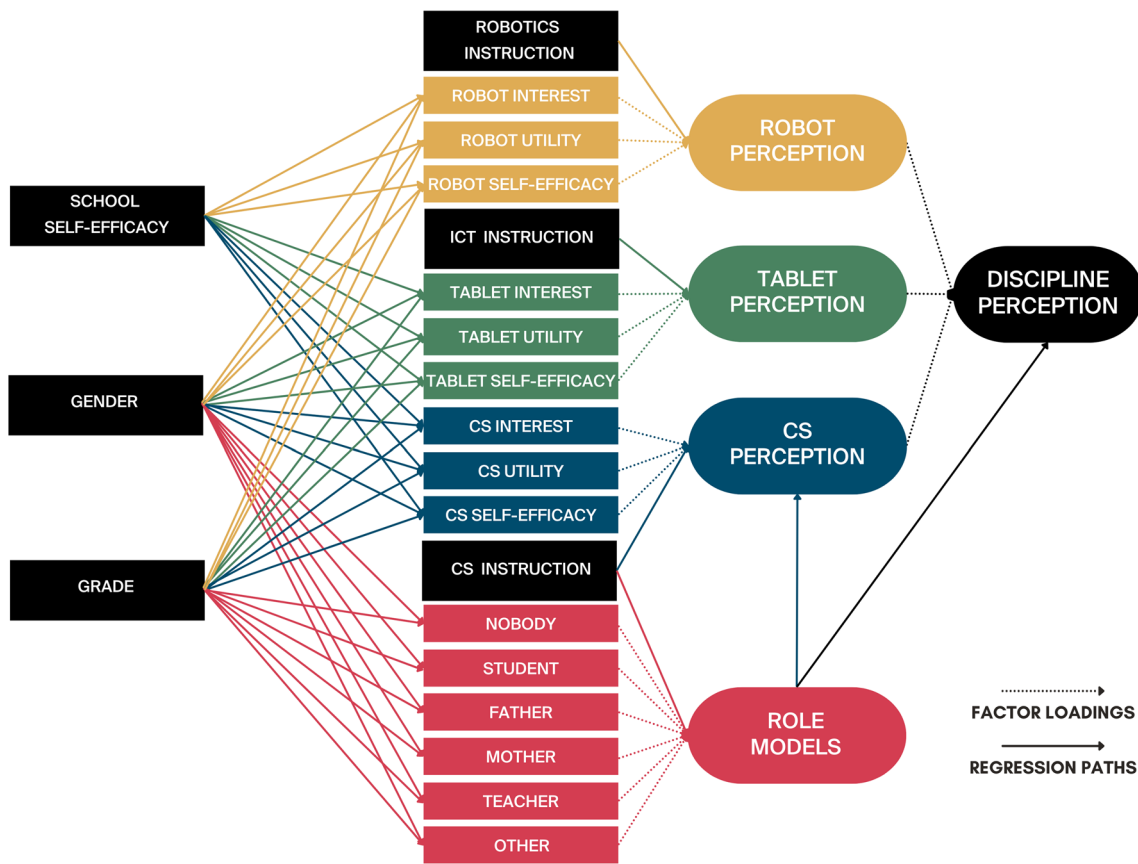
Fig. 7 Structural Equation Model for the perception survey (study 2)

Fig. 8 Structural Equation Model for the link between perception and performance (study 2). Performance here is measured with the cCTt which targets CT-concepts which align with a subset of the CS concepts in the curriculum, i.e. sequences, loops, if-else statements, while statements). Please note that this model includes all paths from the model in Fig. 7 but has been simplified for visualisation purposes

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 20 of 43

Fig. 9 Students' perception in schools that had been teaching CS for three years (n = 2433, November 2021)

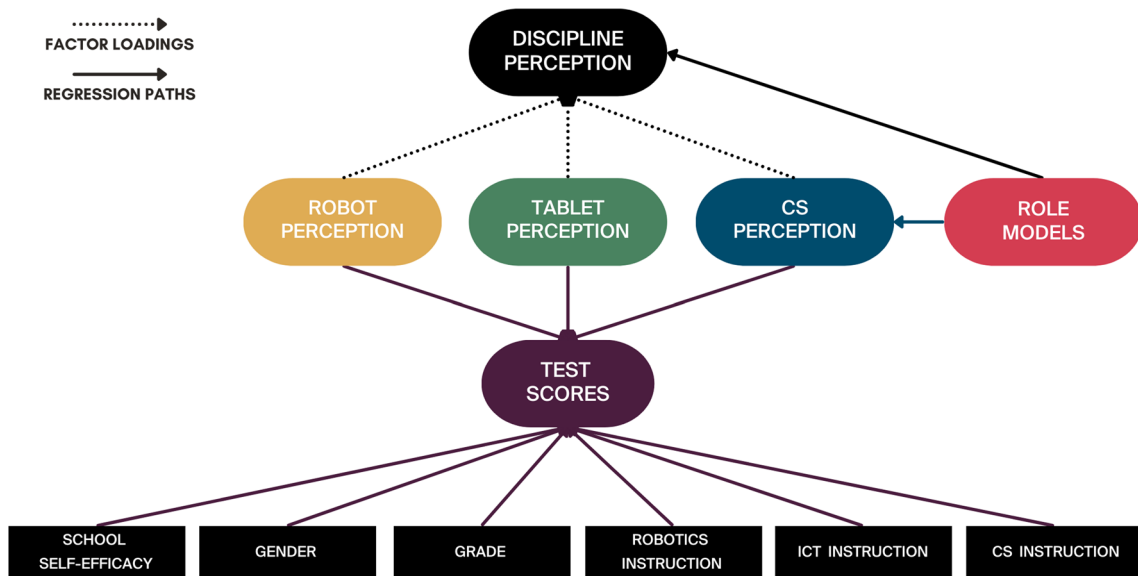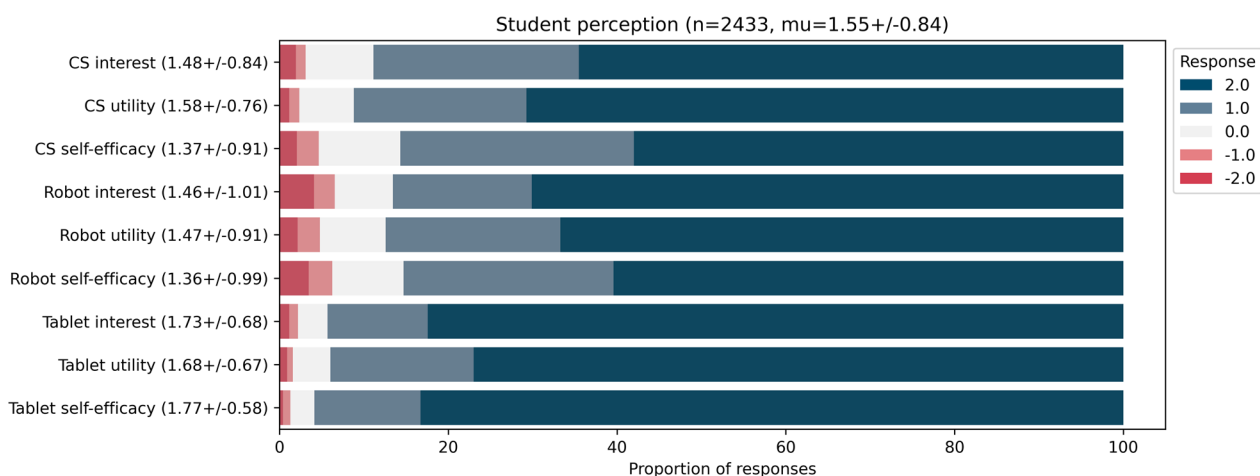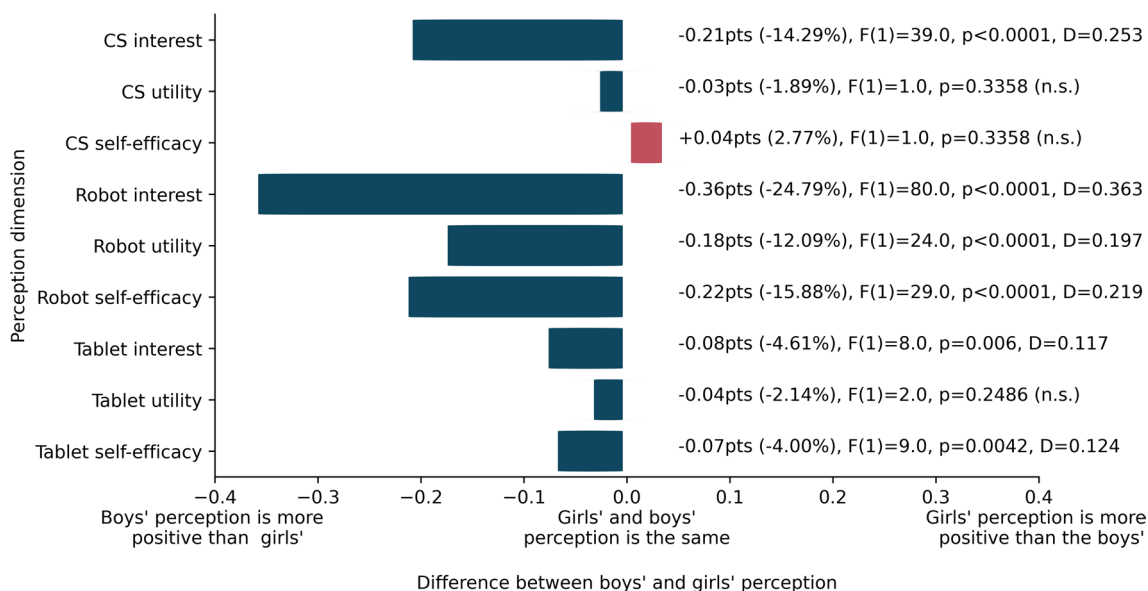**Fig. 10** Delta between boys' and girls' perception on the 5-point Analogue Visual Scale (scores between -2 and $+2$) in schools that had been teaching CS for three years (n = 2433)

- First, a CFA indicates that the measurement model does not have an adequate fit with the modification indices indicating that the issue is due to the "other" option in the CS role model question (Bartlett's test of sphericity $\chi^2(105) = 3660$, $p < 0.001$, KMO = 0.70, model fit $\chi^2(84) = 373$, $p < 0.001$, $\chi^2/df = 4.44$, CFI = 0.888, TLI = 0.859, RMSEA = 0.040, RMSEA $0.90ci = [0.036, 0.045]$, SRMR = 0.039).

- Second, a CFA conducted after removing the "other" option from the CS role model question indicates that the measurement model has an adequate fit (Bartlett's test of sphericity $\chi^2(91) = 3421$,

$p < 0.001$, KMO = 0.72, model fit $\chi^2(71) = 216$, $p < 0.001$, $\chi^2/df = 3.05$, CFI = 0.939, TLI = 0.922, RMSEA = 0.031, RMSEA $0.90ci = [0.026, 0.036]$, SRMR = 0.033).

- Finally, employing SEM on the model in Fig 7 (see "Analysis methodology (study 2)") meets the fit requirements (Bartlett's test of sphericity $\chi^2(190) = 7300$, $p < .001$, KMO = 0.72, model fit $\chi^2(113) = 260$, $p < 0.001$, $\chi^2/df = 2.30$, CFI = 0.941, TLI = 0.908, RMSEA = 0.025, RMSEA $0.90ci = [0.021, 0.029]$, SRMR = 0.026).

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60
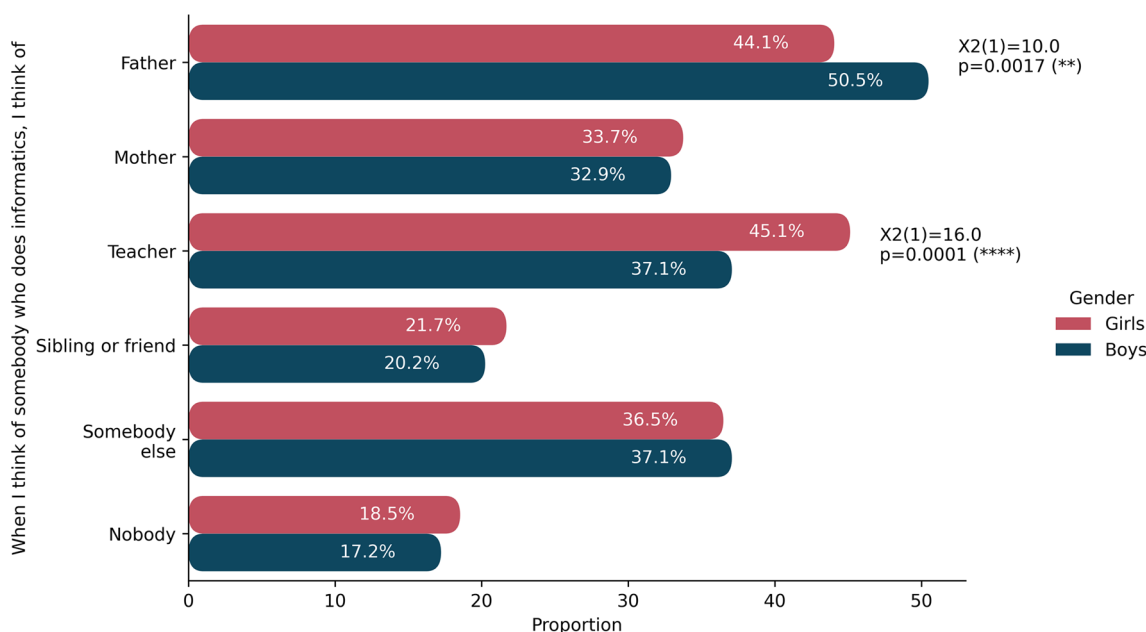
Page 21 of 43

Fig. 11 Students' perception of who does CS in schools that had been teaching CS for three years ($n = 2433$)

Figure 12 shows the significant paths and factors in the model (see Table 13 in Appendix B.1 for all links) and indicates that:

- Perceiving an influencer or somebody close as doing CS (e.g. teacher: $\beta = 0.17$, $p < 0.001$; parent: $\beta_{father} = 0.3$, $\beta_{mother} = 0.23$, $p < 0.001$; or peer: $\beta = 0.13$, < 0.001) positively contributes to the perception of role models, while perceiving nobody has a negative influence $\beta = -0.024$, $p < 0.001$). The role model latent factor then impacts the perception of CS ($\beta = 0.3$, $p = 0.016$) and of the discipline overall, i.e. the second order latent factor in the SEM, $\beta = 0.15$, $p = 0.003$).
- Higher school-related self-efficacy positively correlates with the perception of the discipline on all the Likert scale CS, robot and tablet-related criteria, with the exception of interest in tablets.
- Girls tend to have a more negative perception of the discipline with respect to robots overall, tablets and CS interest, and tablets self-efficacy. Compared to boys, they also perceive the father less often ($\beta = -0.06$, $p = 0.005$) and the teacher more often as doing CS ($\beta = 0.06$, $p = 0.003$).
- Older students are more likely to consider CS ($\beta = 0.09$, $p < 0.001$), tablets ($\beta = 0.03$, $p = 0.020$) and robots ($\beta = 0.07$, $p = 0.000$) useful; while being less interested in tablets ($\beta = -0.04$, $p = 0.014$). They are also less likely to perceive their teacher ($\beta = -0.04$, $p < 0.001$), mother ($\beta = -0.06$,

$p < 0.001$), and nobody as doing CS ($\beta = -0.03$, $p < 0.001$).
- The amount of CS education received since the start of the year does not significantly influence student perception on any dimensions ($p > 0.05$).

The lack of influence between teachers' adoption of CS pedagogical content and perception appears conjointly with a lack of influence between perception and performance. Indeed, the SEM that includes students' scores ($n = 1583$, see Fig. 8 in "Analysis methodology (study 2)") to see how performance is influenced by perception and demographics indicates that there is no significant link (see Table 7). The only variables that significantly influence the score are the grade (older students have higher scores) and their general self-efficacy (students that are more confident in their capacity to succeed in school have higher scores).

Synthesis and limitations of study 2

The students have a positive perception of the discipline, and the tools employed to teach it in schools with access to CS education. Although the results are nearly saturated, the structural equation models help identify that:

- Gender influences the way the discipline is perceived, as girls have a more negative perception of the discipline then boys (in particular where robotics is concerned) which is aligned with stereotypes in these fields (social barriers).
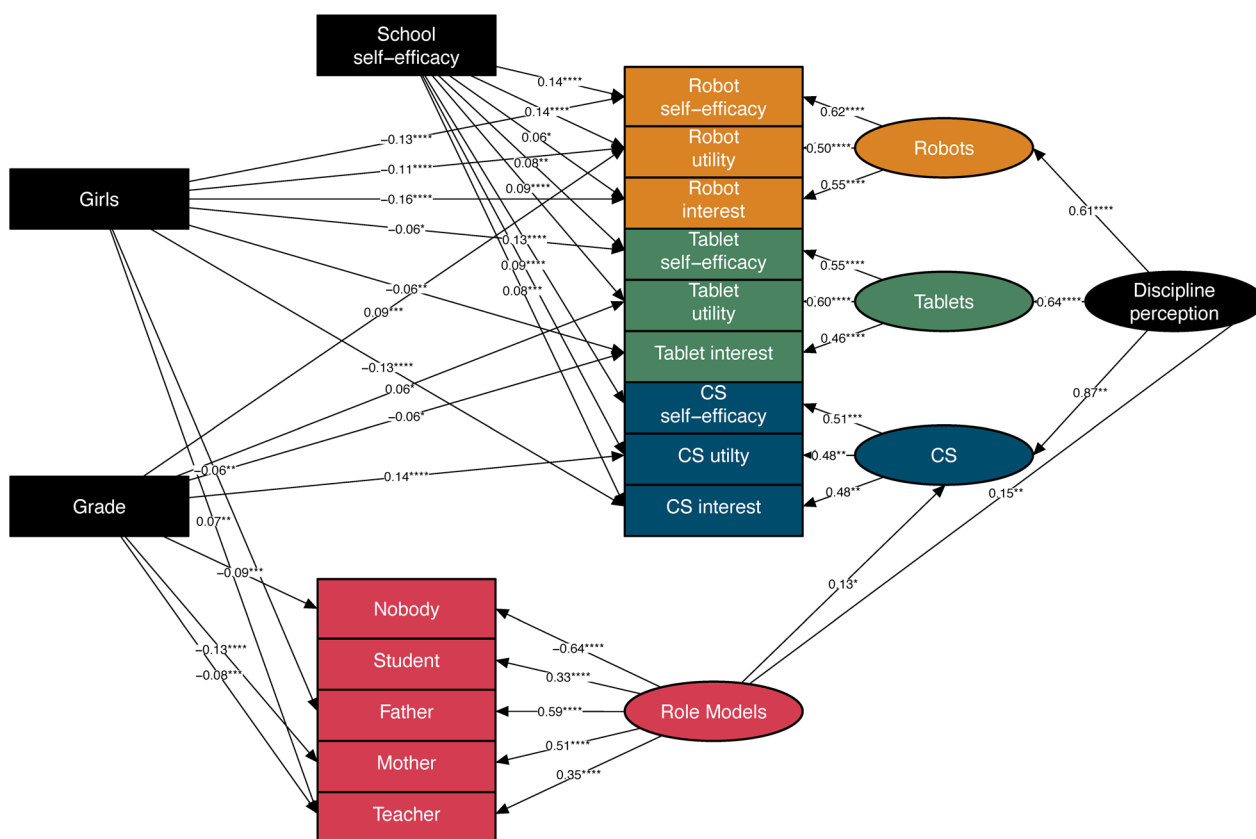
El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 22 of 43

Fig. 12 Perception SEM (*n* = 2116, November 2021) path diagram with standardised variables for the measurement model that meets the requirements for adequate fit displaying only significant links in the model. Please note that all standardised factor loadings are above 0.3

- Having a role model close to the students as doing CS positively influences the perception of CS and the overall discipline, but those perceived as doing CS differs according to gender since girls perceive the teacher more often, and boys the father more often as doing CS.
- There is no influence of the CS education received from the start of the year on perception.
- There is no link between students' perception of the discipline and their performance on the assessment (positive for equity).
- Students' general school-related self-efficacy positively correlates with the perception of the discipline and with students' performance on the cCTt.

There are, however, several limitations to this study, mainly that (i) the students were at least in their third year of CS education by the time the study was conducted, (ii) their perception was positively saturated and (iii) there was no control group. It would have been interesting to have access to a pre-test prior to their first CS lecture and to compare the evolution of perception over time. Where the link between perception

and what the teachers taught is concerned, as for study 1, some of the findings may be biased by the fact that teachers may be teaching CS pedagogical content that was not included in the PD program and are not accounted for in the analyses.

In terms of the perception survey itself, while the CFA analysis indicates that the perception survey is a short and valid instrument that can be employed to measure grades 3–6 students' perception of the discipline and the tools used to teach it, this is not without its limitations. Indeed, the survey measures interest, utility and self-efficacy concepts with only one item for each dimension (CS, robotics, tablets). Ideally, for each concept and dimension, there would be at least 3–4 items (for interest, utility, and self-efficacy) in order to improve the reliability of the instrument. This owes to our requirement of being able to administer the CS perception survey to grades 3–6 students before the cCTt (and not after to avoid having their performance bias their perception), without taking too much in-class time for both (i.e. the perception survey had to be short and take less than 20 min overall with grade 3 students). Nonetheless, researchers have investigated the reliability of single-item items and have

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 23 of 43

Table 7 Unstandardised regression parameters for the perception and background to performance SEM (n=1583, November 2021, $\chi^2(124) = 221.462, p < 0.001, chi^2/\text{df} = 1.79, \text{CFI} = 0.951, \text{TLI} = 0.923, \text{RMSEA} = 0.022, 90\%ci = [0.017, 0.027], \text{SRMR} = 0.026$)

	Model				
	Estimate	Std. Err.	Z	p	R^2
Percentage (/100)					0.136
CS perception	−0.29	1.31	−0.22	0.824	0.707
Tablets perception	−0.04	0.99	−0.04	0.969	0.372
Robots perception	0.96	1.11	0.86	0.387	0.411
General self-efficacy	**1.54**	0.69	2.22	**0.027**	
Gender (0 = boys, 1 = girls)	−1.57	1.06	−1.47	0.141	
Grade	**7.53**	0.53	14.19	**0.000**	
Number of CS education periods SI	0.09	0.13	0.70	0.483	
Number of ICT education periods	−0.02	0.08	−0.20	0.842	
Number of Robotics education periods	−0.29	0.42	−0.68	0.496	

Please note that on the smaller sample CS utility did not correlate highly with interest and self-efficacy and had to be removed from the model. For the full set of model parameters see Table 14 in Appendix B.2

Table 8 Number of participants in the second student perception survey (study 3, May 2021)

CS education	Gender	Grade				Total
		3	4	5	6	
False (i.e. non-CS-school)	Boys	84	91	116	122	413
	Girls	83	98	121	98	400
	Total	167	189	237	220	813
True (i.e. CS-school)	Boys	102	91	96	135	424
	Girls	93	99	92	123	407
	Total	195	190	188	258	831
Total	Boys	186	182	212	257	837
	Girls	176	197	213	221	807
	Total	362	379	425	478	1644

shown that it is possible to have reliable measures with only single items (see Hoeppner et al. 2011).

Study 3: student perception between CS-schools and schools where teachers were not yet trained to teach computer science

Methodology (study 3)

Participants and data collection (study 3)

To extend study 2, the perception survey (see Table 6) was administered to all students in grades 3–6 ($n = 1644$, see Table 8) from 3 schools with access to CS education (which we refer to as CS-schools, $n \sim= 831$) and 2 similar schools without access to CS education (which we refer to as non-CS-schools, $n \sim= 813$) which were selected to be representative of the demographics of the region. Non-CS-schools here therefore refer to schools where, at the time of the study, the teachers were neither trained to introduce the new discipline into their practice nor had access to the material resources, infrastructure or support they require to teach the discipline, an element which was confirmed by an accompanying teacher survey. The objective was to compare the students' perception of the discipline between the two conditions (CS-schools and non-CS-schools) as students in CS-schools had been in contact with the discipline for multiple years and perception was positively saturated in study 2.

Analysis methodology (study 3)

The comparison between both groups is established using Structural Equation Modelling by constraining the models to have equal factor loadings, and allowing the regression parameters to vary between the two groups (gender, grade, general self-efficacy). By comparing the intercepts of the two SEMs, it is possible to establish the effect of having received several years of CS-education on perception. By comparing the regression parameters, it is possible to establish whether there are interaction effects between the student variables (e.g. gender) and access to

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 24 of 43

CS-education, and thus determine if gender-related gaps are indeed closing with the introduction of the novel curriculum.

Results: perception and the influence of having access to CS education on perception (study 3)

The SEM to compare the groups (CS-schools vs. non-CS-schools) constrained the loadings and thresholds, while leaving the intercepts and regression parameters free to vary between groups[9]. We thus compare the intercepts and regression coefficients between the groups (for the full SEM, see Table 15 in Appendix C.1).

The intercepts for both groups indicate that students' responses positively saturate for both groups for nearly all CS, robotic and tablet perception items (see Fig. 13). Nonetheless, students in CS-schools appear more interested generally, and evaluate the robotics generally more favourably. However, CS and tablet utility and self-efficacy are lower for students in CS-schools. Students in CS-schools perceive the teacher more often as doing CS, which is coherent with the fact that their teachers over the past few years have been teaching CS pedagogical content. On the other hand, students in CS-schools perceive their mothers and other students less often as doing CS, possibly indicating that the students have a better awareness of what it means to "do" CS (Pantic et al., 2018), and that it is not only related to using a computer or tablet.

The significant impact of general self-efficacy and gender on student perception is shown in Figs. 14, and 15.

Figure 14 shows that general school-related self-efficacy positively influences CS self-efficacy ($b_{CS} = 0.16$, $p_{CS} = 0.001$, $b_{no-CS} = 0.2$, $p_{no-CS} < 0.001$) and robotics self-efficacy ($b_{CS} = 0.11$, $p_{CS} = 0.033$, $b_{no-CS} = 0.14$, $p_{no-CS} = 0.016$) of all students. This reveals that students who consider themselves less capable of doing well in schools also think that they are less able to do CS and robotics, although the influence is less pronounced when students have received CS-education. Access to CS education may thus contribute to a wider range of students considering that they are capable of doing CS and robotics. On the other hand, for tablets, while there is no significant influence of school-related self-efficacy in non-CS-schools ($p_{no-CS} = 0.054$), it is present in CS-schools ($b_{CS} = 0.08$, $p_{CS} = 0.016$) which may indicate that students realise the range of possibilities (beyond merely passive activities) and that this may require more

competencies to be able to make use of. Nonetheless, general self-efficacy does not influence interest or perceived utility in CS-schools ($p_{CS} > 0.05$), contrary to non-CS-schools for CS interest ($b_{no-CS} = 0.1$, $p_{no-CS} = 0.036$), CS utility ($b_{no-CS} = 0.14$, $p_{no-CS} = 0.001$), and Robotics' utility ($b_{no-CS} = 0.11$, $p_{no-CS} = 0.044$). It would thus appear that access to CS-education helps reduce these biases.

Where gender is concerned (see Fig. 15), all gender gaps that are identified as significant confirm the stereotypes that boys perceive the discipline more favourably than girls. Some gender gaps are only present in CS-schools (CS and tablet interest and self-efficacy, robots utility) suggesting that access to CS-education increases these gaps which, interestingly, are initially the smaller or non-significant gaps in non-CS schools. There are nonetheless some gaps that are smaller in CS-schools, all the while remaining present in both types of schools: robotics interest and self-efficacy, as well as perceiving the teacher as doing CS in CS-schools, which interestingly are the initially larger gaps in non-CS schools. Only the CS-interest gap is present in both schools and stronger in CS-schools.

Synthesis and limitations of study 3

Students' perception of the discipline is highly positive and affected by gender biases (social barriers) in both schools with CS education and schools without. However, access to CS education leads to:

- Positive impacts through: increased interest in CS and the associated tools, a more positive perception of robotics on all dimensions, teachers being more often perceived as doing CS.
- Negative impacts through: lower self-efficacy with respect to CS and tablets.
- Positive outcomes for equity through: closing larger gender-gaps in terms of robotics interest and self-efficacy (gender-equity), a lesser influence of general self-efficacy on several perception dimensions (CS interest, utility, self-efficacy; robotics utility and self-efficacy).
- Negative outcomes for equity through: increasing smaller gender-gaps in terms of CS and tablets self-efficacy (gender-equity), a higher influence of general self-efficacy on tablets' self-efficacy.

As in the case of studies 1 and 2, this study has its limitations. Firstly, the sample is relatively small to do a comparison between groups (even when constraining parameters to be equal). As such, the minimum effect size that can be detected is smaller than in the case of study 2. This analysis would therefore benefit from a replication at a larger scale. As mentioned for study 2, there

---

[9] The selection of model constraints was achieved by successively comparing through ANOVA the following SEMs: (1) without groupings, (2) groupings without constraints, (3) constrained loadings and thresholds, (4) constrained loadings, thresholds, regression parameters, (5) constrained loadings, thresholds, intercepts.

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 25 of 43
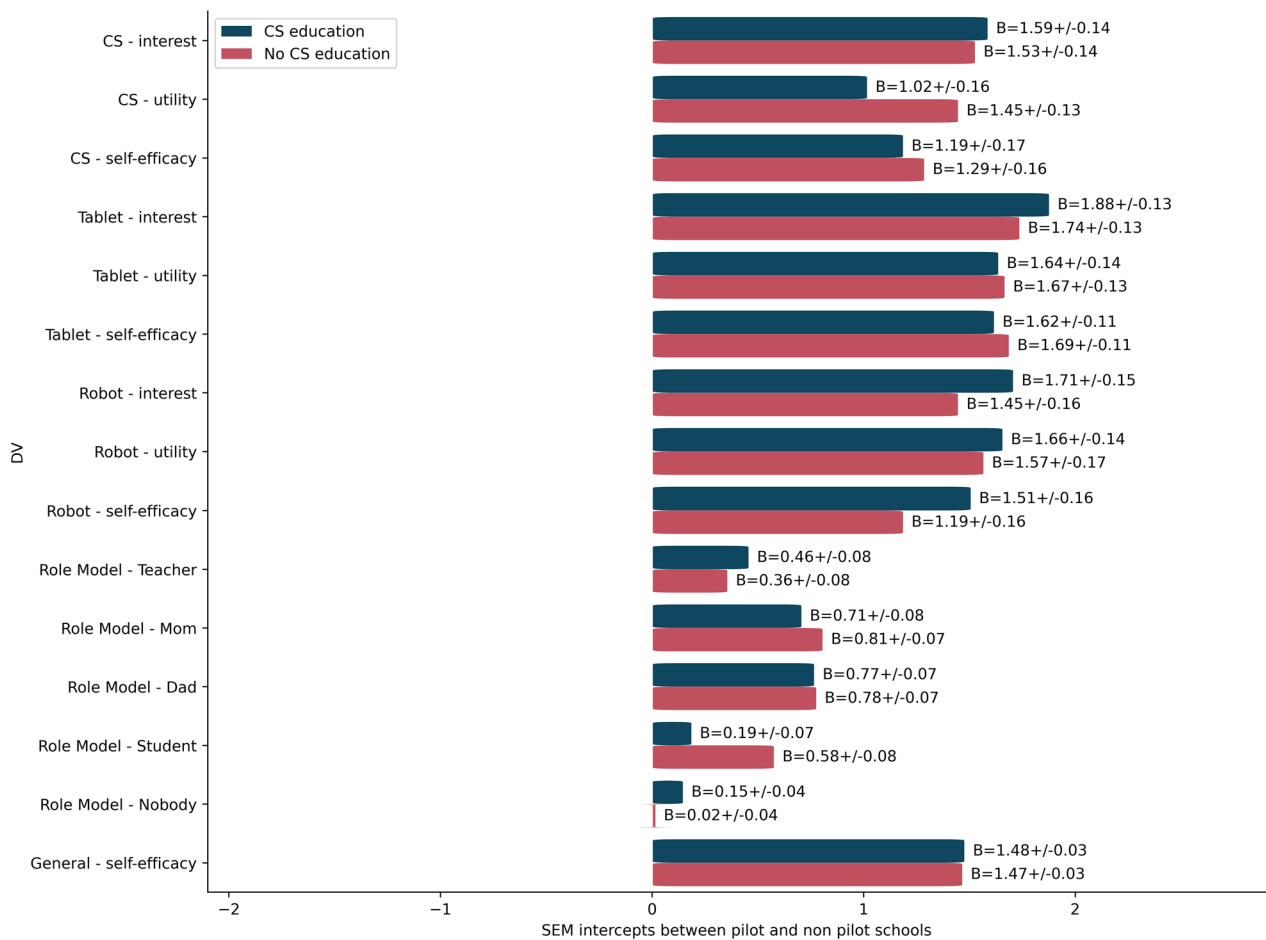
Fig. 13 Comparison of the SEM intercepts between schools that had access to CS-education and schools that did not

Fig. 14 Comparison of the SEM regression coefficients for general self-efficacy between schools that had access to CS-education and schools that did not. Please note that only significant regressors are shown

is also no view on how the perception evolves over time within these groups, and at the point where students gain access to CS education the first time. Therefore, it would be interesting to have access to a sample of students just before they began having access to CS education and then follow up over time, and compare with a group that has no access to CS education. This type of analysis has temporal constraints and must be planned for at the start

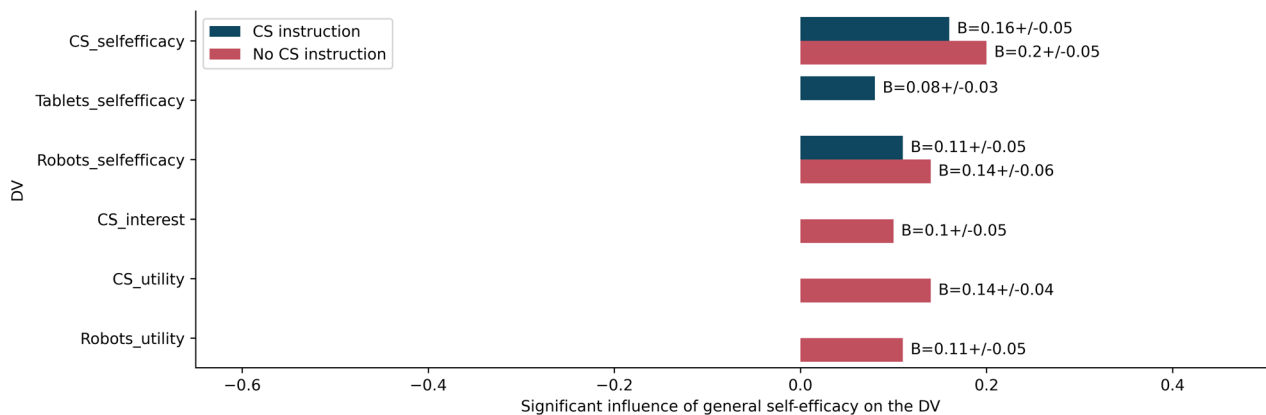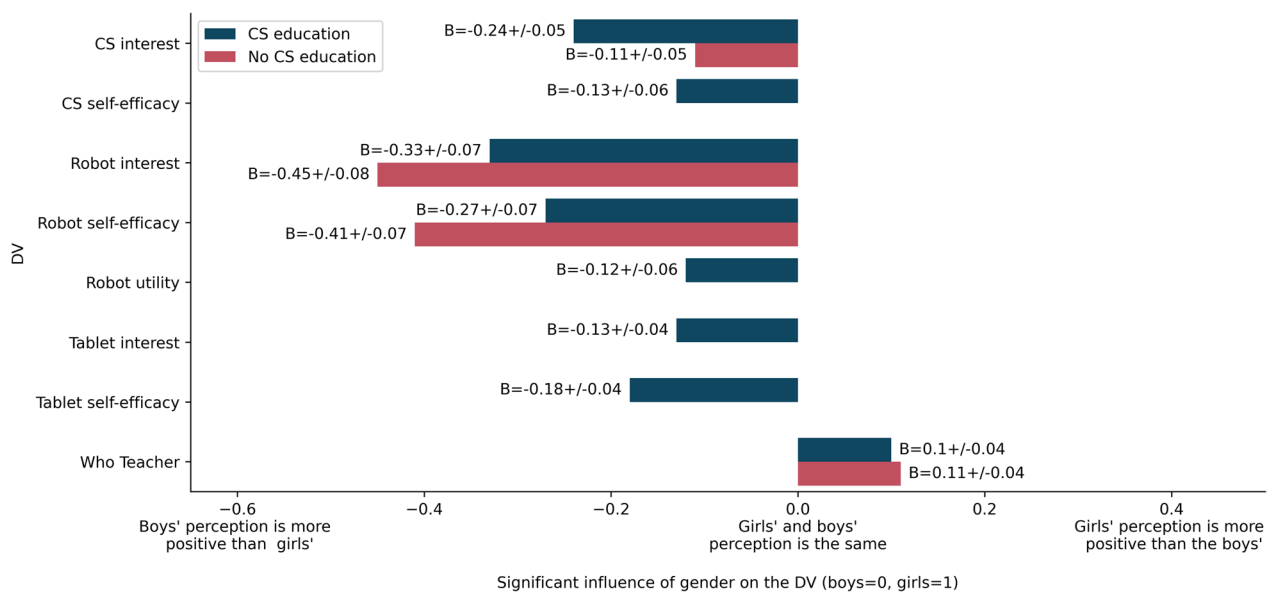El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 26 of 43

Fig. 15 Comparison of the SEM regression coefficients for gender between schools that had access to CS-education and schools that did not. Please note that only significant regressors are shown

of the reforms and prior to deployment to all schools if the objective is to be able to compare for an extended period of time.

Discussion

This article investigates whether a large-scale mandatory primary school CS curricular reform and accompanying PD programme has an impact, and contributes to achieving equity goals, in terms of learning and perception. As indicated in the introduction, achieving equity goals requires addressing structural (i.e. access related) and social (i.e. stereotype related) barriers that lead to under-representation in the field by influencing performance and perception early on. While equity in terms of access is ensured by the fact that the reform is being deployed to all teachers in the region, two main questions drive the study:

(RQ1) How does teaching CS pedagogical content impact student learning? And how does it impact learning-related gender- and performance-equity?

(RQ2) How does teaching CS pedagogical content impact students' perception of CS? And how does it impact perception-related self-efficacy- and gender-equity?

We provide a visual synthesis of the findings in Fig. 16 based on the learning and perception data drawn from 3 studies that were conducted over 2 years and involved, respectively, $n_1 = 1384$, $n_2 = 2433$ and $n_3 = 1644$ grade 3–6 students (ages 7–11) and their $n_1 = 83$, $n_2 = 142$ and

$n_3 = 95$ teachers. The findings are further discussed in the following subsections.

Impact of the curriculum reform on student learning, and learning-related performance- and gender-equity (RQ1)

Student learning impact

The findings of studies 1 and 2 indicate that the students progress in terms of CT-concepts (sequences, loops, if-else statements, while statements) over time, consistently with other studies that have found that students' algorithmic skills improve as they age (Kanaki & Kalogiannakis, 2022; Piatti et al., 2022). In particular, we observe that grade 3 students achieved a year's worth of CT-development in the 6 months that separated the pre- and post-tests (positive impact). Indeed, the grade 3 students' post-test scores were equivalent to the grade 4 students' pre-test scores. However, there is no direct link between learning and the amount of CS education received (absence of impact). There is, on the other hand, a positive influence of the teachers' perception of the PD programme (positive impact), which may be due to the Pygmalion (or Rosenthal) effect according to which a teachers' expectations may act "as a self-fulfilling prophecy" (Rosenthal, 2010). While teachers' perception of the PD-programme likely acts as a mediating variable for teachers' assimilation of the underlying CS-concepts and their appropriation of the pedagogical content, it does indicate the need to find means of motivating teachers to introduce CS into their practices (El-Hamamsy et al.,
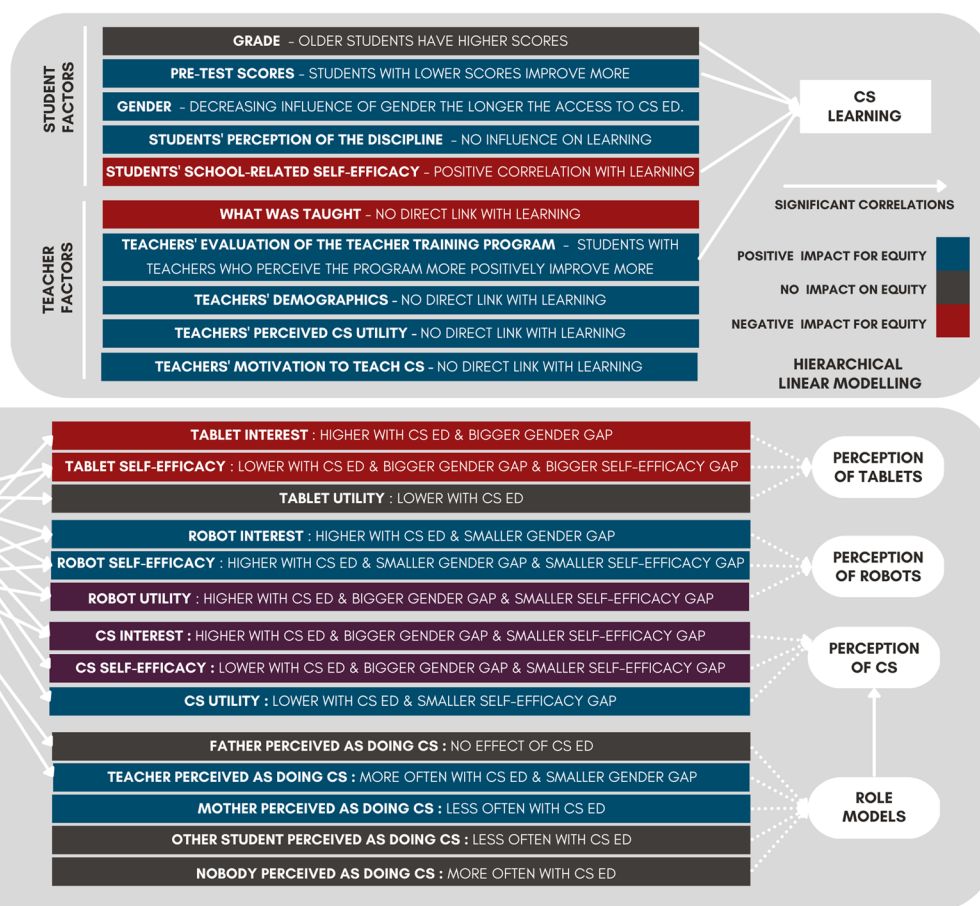
Fig. 16 Visual synthesis of the study's findings and how these relate to impact and equity. Each factor considered is indicated in a rectangle in bold. For the student learning results that are based on hierarchical linear models, the identified effect of said factor on the outcome variable is indicated in plain text in the same rectangle. For the student perception results that are based on structural equation modelling, the impact of general school-related self-efficacy and gender on a given factor are indicated in the factor's rectangle in plain text. In both cases, the impact of the measured effect (or lack thereof) on equity is colour coded (blue for a positive impact, red for a negative impact, purple for a mixed impact and black for an absence of impact). Please note that we only indicate significant links/effects (i.e. $p > 0.05$) which does not reflect on the strength of the effect detected (for that please refer to the results section and see the effect sizes and regression coefficients)

2022a) and to ensure that they see the utility of doing so (El-Hamamsy et al., 2023b).

The lack of a direct link between what the teachers taught (i.e. adoption) and learning could be due to two main factors and their interaction: the adequacy of the content with respect to the targeted concepts, and the teachers' appropriation of the CS pedagogical content. We have synthesised the corresponding hypotheses in Table 9 depending on whether either or both of these factors are indeed at play in the present context. As a reminder: the teachers were trained to introduce the specific CS-pedagogical activities which were designed by experts in CS and pedagogy from multiple institutions. Therefore, considering conjointly these elements, and the link between student learning and the teachers' perception of the PD-program, it appears likely that the second hypothesis is true. More specifically: the lack of

direct link with adoption could be partially or entirely due to teacher-level factors (their mastery of the concepts, and how they are teaching the pedagogical activities), although we may not presently rule out the other hypotheses.

To better understand the impact of teaching CS on learning, it would be important to investigate the various hypotheses by considering teacher assessments to gain insight into their mastery of the concepts, classroom observations to gain insight into teachers' implementation fidelity, and comparing with students in non-CS-schools. Such an approach would not only make it possible to assess each of the pedagogical activities individually, but would also give the opportunity to provide guidelines regarding how best to teach the pedagogical content to promote learning. Doing so, however, requires getting past certain barriers in the field, whether in terms

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 28 of 43

Table 9 Hypotheses related to the absence of direct links between CS-education and student learning

		The teachers' appropriation of the CS-pedagogical content is aligned with the curricular objectives	
		True	**False**
The CS-pedagogical content is adequate with respect to the targeted concepts	True	H1: The students have reached the limit of their cognitive abilities and are not capable of progressing more, irrespective of the additional content and CS education received	H3: The teacher, while teaching the CS-pedagogical activities is not teaching the CS-concepts well (H3.1) either because they do not have sufficient mastery the concepts themselves; or (H3.2) because they do not put the emphasis on the CS concepts while teaching and focus on other facets, such as disciplinary links (e.g. maths or verbalisation), coherently with the differences between intended, enacted and attained curricula that are present generally van den Akker (2003) and in the context of CS Falkner et al. (2019). The PD program should be revised.
	False	H2: The CS-pedagogical content is either (H2.1) not developmentally appropriate Ottenbreit-Leftwich & Yadav (2022); Bers et al. (2022b), or (H2.2) does not go sufficiently in depth for students to progress beyond what they are acquiring without the CS-education, and should be revised.	H2 + H3

of teacher reticence towards classroom observation and evaluation (Hickmott & Prieto-Rodriguez, 2018), or in terms of acceptability for policy and decision-makers (e.g. access to a control group for performance assessments). However, it is only by gaining such insight that it will be possible to adapt the CS curricular reforms so that it is successfully implemented and sustained in teachers' practices. This could include adapting the CS pedagogical content, PD program, or even considering a different strategy to introduce CS into formal K-12 education. The latter could involve having specialised teachers, or introducing CS transversally to support other disciplines thus contributing to "build computational literacies in all students" (Peel et al., 2022), all the while accounting for time struggles (Ottenbreit-Leftwich & Yadav, 2022) which according to Fofang et al. (2020) would provide pedagogical and equity benefits (but may also run the risk of decreasing the impact of the curricular reform, Suessenbach et al., 2022). We therefore argue that a complete assessment of CS and CT curricula would benefit greatly from expanding to other dimensions of CT (e.g. CT-processes, Brennan & Resnick, 2012), and evaluating the impact that CS-pedagogical content may have on learning in other disciplines (El-Hamamsy et al., 2022b; Ottenbreit-Leftwich & Yadav, 2022), transversal and twenty-first century skills (Barr et al., 2011; El-Hamamsy et al., 2022b; Gretter & Yadav, 2016; Nouri et al., 2020).

Student learning equity

The findings of study 1 indicate that students performing lower at the pre-test progress more in the 6 months before the post-test. This indicates that the performance gap is closing and contributing to *performance-equity* and is consistent with Vygotsky and Cole (1978)'s concept of the Zone of Proximal Development (ZPD). The ZPD is determined by the learning activity and its relation to what students are capable of doing along and with a specific instruction. Therefore, it would appear that the content is adapted to all students because:

- Students with low scores on the pre-test progress more, indicating that the pedagogical content is within their ZPD.
- Students with high scores on the pre-test may already master the concepts and therefore not progress more with the instruction provided.

Provided the additional lack of influence of teacher-demographics (including ICT experience, teaching experience and age which have been found to impact student achievement in various contexts Burroughs et al., 2019; Croninger et al., 2007; Kini & Podolsky, 2016; Ladd & Sorensen, 2017), their perceived utility of CS and their autonomous motivation to teach CS, on student learning, this would appear to indicate that the PD-program contributes to fostering student learning, and *learning equity* more generally.

The findings of study 1 also indicate that a marginally significant gender gap exists in grades 3–4 (likely due to stereotypes and social barriers), and that it appears to be closing over time (positive for gender-equity). This is corroborated by the data from study 2 (from the following academic year) where students who have had more access to CS education do not exhibit gender gaps. These

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 29 of 43

findings therefore confirm the importance of providing prior CS experience to address performance-related gender gaps. As the study did not include grade 1–2 students, it would appear relevant to follow up on the cohort of students over multiple years (and from the start of their schooling) to see how these differences appear and evolve over time.

The findings therefore appear to support that the CS-curricular reform contributes to achieving *learning equity goals.* This would align with the findings of a recent independent study conducted in Germany to evaluate the impact of the introduction of "informatics" into the curriculum throughout the country. In a longitudinal study, Suessenbach et al. (2022) found that (i) lower secondary schools students' ICT competence increased with access to informatics education; (ii) the gap between students with low and high socioeconomic backgrounds decreased; (iii) gender gaps were closing with girls catching up with boys' performance; and (iv) the impact was stronger in the case of informatics as its own discipline rather than having informatics transversally integrated into other subjects.

Impact of the curricular reform on student perception and perception-related self-efficacy- and gender-equity (RQ2)

Student perception impact
Students' perception of the discipline and the tools employed to teach it is globally positive in primary school, whether in CS schools or not (studies 2, 3), as the results are positively saturated. Nonetheless, students' overall perception of the discipline is influenced by access to CS education. Indeed, access to CS education contributes to increased interest in CS and the associated tools, with a more positive perception of robotics overall (positive impact). Perceived utility and self-efficacy towards CS and tablets are however lower (negative impact). The latter may be indicative of a better understanding of what CS is, and the extent of the applications that are possible with tablets, contributing to more realistic expectations (Pantic et al., 2018), possibly addressing a key issue identified when introducing CS education in secondary school (El-Hamamsy et al., 2023c). As the results remain globally positive, they appear promising for both CS and robotics, particularly since interest, self-efficacy and perceived utility are key motivational factors that influence academic performance and career choices. Future studies should therefore i) continue to monitor how these factors evolve and how they relate to students' decision or not to pursue studies in these fields (which in the present educational system, begins at the end of 8th grade) and ii) investigate using qualitative methodologies why certain trends are observed.

Student perception equity with respect to the effect of gender
Gender gaps are present already in grades 3-6 with boys having a more positive perception of the discipline than girls on nearly all criteria, coherently with Master et al. (2021)'s and Sullivan and Bers (2016)'s findings, and despite access to CS-education from grade 1 (study 2). Robotics in particular appears to be subject to the largest gender gaps (study 2, 3). Nonetheless, the perception of CS-role models, and in particular influencers (Wang & Hejazi Moghadam, 2017) such as teachers and parents being perceived as doing CS has a positive influence on the perception of the discipline, but is subject to gender biases (study 2). As access to CS-education contributes to more students perceiving their teachers as doing CS (study 3), and these teachers are mainly women in primary school, the introduction of CS for all in schools may contribute to addressing social perceptions and counter the creation of early gender gaps evoked by Wang and Hejazi Moghadam (2017). The model selection further indicates that the influence of gender on perception varies with access to CS education (impact on gender-equity). While for interest and self-efficacy the gender gap appears to be closing for Robotics in CS-schools (positive for gender-equity), the gap is increasing for CS and tablets (negative for gender equity). Different trends along these dimensions indicate that the CS pedagogical activities might need to be adjusted to "provid[e] students with early experiences that signal equally to both girls and boys that they belong and can succeed" (Cheryan et al., 2017) (e.g. by adopting more collaborative settings, Sullivan and Bers, 2016, or introducing social aspects). The findings further indicate that introducing robotics as a means of teaching CS may also contribute to broadening participation in STEM by reducing robotics-related gender biases. This complements a prior study that found that employing robots to teach CS benefited both CS and robotics at the teacher and PD-level (El-Hamamsy et al., 2021a). As such, robotics to teach CS benefits both the PD-, teacher-, and student-levels. Robotics and STEM more broadly could therefore take advantage of ongoing CS-curricular reforms worldwide to broaden participation and engage more students in these fields.

Student perception equity with respect to the effect of school-related self-efficacy
The influence of general self-efficacy varies between CS and non-CS schools (study 3). On the one hand, in CS-schools, there is a positive influence of general self-efficacy on tablet-related self-efficacy, which is not the case for non-CS-schools. Once again, this may be due to students having a better awareness of what it means to "do" CS and more creative activities with tablets (Pantic et al., 2018). Indeed, the tablet is a ubiquitous tool

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 30 of 43

in the region which students easily have access to. Their traditional usage of this tool differs significantly from the type of activities that are proposed in the curricular reform which tend to be active and creative. The learning objectives of these tasks push students to adjust their "imagination" around this tool (Flichy, 2001). Concretely, we believe that the students are forced to reconsider the affordances and the potential of this tool, thus reevaluating their own competencies beyond the more traditional use involving playing games, texting, taking photos and watching videos.

On the other hand, CS perception (interest, self-efficacy and utility), and robots perception (self-efficacy, and utility) are positively influenced by general self-efficacy in both CS and non-CS schools, but to a lesser degree when students have received CS education. Contrary to tablets, CS and robotics are novel, with students having little to no access in schools (or at home) where the CS education curricular reform has not yet occurred. Therefore, we believe that the positive influence of general self-efficacy on domain-specific self-efficacy is consistent with Bandura (1986)'s sociocognitive theory on auto-evaluation: people's belief in their efficacy to do a task is developed through vicarious experience, i.e. by comparing themselves to others. However, it is also built through mastery experiences: by experiencing CS and robotics-related activities, the students are more influenced by their own CS- and robotics-specific experiences, and less by their overall assessment of their capacity to succeed in school. Therefore, given the influence of self-efficacy on students' choices and career decisions in the long term, such experiences may ultimately contribute to broadening participation in the field to a wider range of students, and namely to not only those who believe they are good in school.

Student perception equity with respect to the link between performance and perception

There is *no evident link between student performance and perception of the discipline* (study 2, positive for equity), such as those found in other studies in middle school (Hinckle et al., 2020); Rachmatullah et al., 2022). However, student performance is related to students' general self-efficacy, with students who consider that they are better at school performing better on the test. This would suggest that there may be a link between students' performance on CT-concepts and other disciplines, irrespective of how students perceive the discipline. This may be indicative that perception is not yet biased by performance and inversely. Nonetheless, given the role that perception (and stereotypes) has been found to play on academic and career decisions (see "Introduction and

related work"), it is important to continue to monitor how students' perception evolves over time and establish at which point this may influence their sense of belonging and career decisions.

Globally, the trends observed confirm not only the importance of introducing the discipline in formal education for all, but also the complex interactions that this introduction may have on students' perception. The latter indeed may not necessarily contribute to closing all perception-related gaps but may also exacerbate others. Therefore, in addition to conducting the study with a larger sample to be able to detect smaller effect sizes, it would be important to complement the results of the study with qualitative data to gain better insight into how students perceive the discipline, how this differs, and why, between students with and without access to CS-education

Conclusion

Early exposure to Computer Science (CS) and Computational Thinking (CT) for all is important to broaden participation and promote equity in the field. This is contingent on addressing structural related barriers (lack of access) and social barriers (stereotypes) in order to reduce performance and perception gaps which affect sense of belonging and career decision. Addressing these barriers requires a system-wide implementation of CS & CT curricula for all students starting early foundational years. That is why numerous countries are introducing CS & CT into their curricula starting primary school. The question is therefore: are these curricular reforms contributing to learning and reducing performance gaps? Curricular reforms and professional development programmes are seldom evaluated at the student-level despite the importance of establishing their effectiveness in terms of student learning and perception. Therefore, in the present article, we evaluate the implementation of a regional CS-curricular reform in order to determine if the reform contributes to achieving equity goals. More specifically, we study how the implementation of the CS curriculum by teachers impacts and contributes to equity in terms of student learning (with respect to gender and performance gaps, RQ1) and perception (with respect to gender and self-efficacy gaps, RQ2). To answer these questions, the analysis employs hierarchical linear modelling and structural equation modelling using data from three studies involving, respectively, $n_1 = 1384$, $n_2 = 2433$ and $n_3 = 1644$ grade 3–6 students (ages 7–11) and their $n_1 = 83$, $n_2 = 142$ and $n_3 = 95$ teachers.

In terms of student learning impact, the students are progressing over time. There is however no direct link between what the teachers taught (i.e. adopted) over an

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 31 of 43

extended period of time and student learning. Although certain studies have suggested that perception may play a mediating role on performance, this is not the case in the present study. There is however a link between student learning and how teachers perceived the CS-PD program. Teacher perception may thus be acting as a mediating variable or be confounding with other dimensions such as teachers' assimilation of Technological Pedagogical and Content Knowledge (Mishra & Koehler, 2006) obtained during the PD, their appropriation of the content and the depth of the associated change in their practice, supporting the need to gain better insight into how the content is taught. As there are known differences between intended, enacted and attained curricula (van den Akker, 2003), the findings indicate the need to investigate not just whether, but how teaching the discipline, and individual pedagogical content, influences learning. In terms of student learning equity, the findings indicating that (i) the performance gap between lower and higher achieving students are closing and that (ii) pre-existing gender gaps appear to be closing. Whether in terms of impact or equity, it would be important to expand to other dimensions of learning that may be influenced, whether to have a more complete evaluation of CT (by including practices and perspectives, Brennan & Resnick, 2012), or by looking more generally into the impact on learning in other disciplines, or in terms of transversal competences.

Where student perception is concerned, in terms of impact, the results are relatively straightforward: students in both CS and non-CS schools perceive CS and the tools involved with teaching CS positively. Interest in the discipline and perception of robotics is nonetheless more positive in schools with access to CS-education which may contribute to broadening participation in the field. The findings in terms of equity indicate that there are gender gaps which indicate that boys have a better perception of the discipline than girls. However, whether in schools with access to CS education or not, the perception of role models close to them as doing CS contributing to student's positive perception of the discipline. As teachers are mainly women in primary school, introducing CS as a discipline taught by all teachers contributes to teachers being more often perceived as doing CS, and may ultimately contribute to gender-equity. Comparing students in schools with and without access to CS education indicates that there are differences in how the discipline is perceived in both types of schools and that there are interaction effects with gender: ii) initially smaller gender gaps are widening (e.g. CS and tablet interest and self-efficacy, robots utility) while initially higher gender gaps are closing (e.g. robotics

interest and self-efficacy, perceiving teachers as doing CS in CS-schools) with access to CS-education. Teaching CS thus has a complex influence on perception which requires investigating more deeply why students perceive the discipline the way they do and how it is influenced by access to CS-education. Monitoring this perception over time is also critical in order to understand how it evolves over time and influences long-term career decisions.

Answering the overarching question "how does the curricular reform impact student-level outcomes and equity in the field?" is therefore not as straightforward as it seems. On the one hand, introducing CS for all in the curriculum and being taught CS has a positive impact and affects equity by:

- Promoting student learning and contributing to performance-equity by reducing (i) differences between initially high and low performing students; (ii) the performance gender gap; and (iii) the impact of teacher demographics on student learning.
- Contributing to perception gender-equity by reducing the largest gender-related perception gaps (namely those pertaining to robotics).

On the other hand, the curricular reform does not automatically lead to improvements on all fronts. The impact is neither direct, as shown by the student learning results which lack a direct link between what was taught and learning; nor straightforward, as shown by the fact that there is an interaction effect between gender and access to CS education, with initially smaller (or not initially present) gender gaps increasing.

The findings of the study therefore demonstrate that the following elements are important to achieve equity and broadening participation in the field:

- Introducing CS for all students starting the first years of formal education.
- Preparing the teachers to teach CS, removing the influence of teacher demographic and teacher motivational factors on student learning.
- Having activities that signal to girls and boys equally and that are in students' Zone of Proximal Development in order to help all achieve the desired learning objectives.
- Investigating the impact of CS curricular reform and PD program implementations at the student level, and including teacher-level insight, all the while considering that the complex dynamics that may be involved in CS-education.

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 32 of 43

Appendix A: Appendix for study 1

A.1 Student demographics for study 1b

Table 10 Student learning demographics for the data subset which considers n=989 students with complete January and June test data and adoption data (study 1b)

	Metric	Grade 3		Grade 4		Total
		No CS education classes	CS-education classes	No CS education	CS-education	
Students	Count	71	409	113	396	989
Classes	Count	4	24	6	21	55
Activities	M±SD	0.00 ± 0.00	3.29 ± 2.02	0.00 ± 0.00	3.12 ± 1.59	
	Range	[0; 0]	[1; 9]	[0; 0]	[1; 6]	
Hours spent	M±SD	0.00 ± 0.00	16.12 ± 11.25	0.00 ± 0.00	12.54 ± 6.80	
	Range	[0; 0]	[1; 48]	[0; 0]	[2; 26]	
January cCTt Score	M±SD	12.37 ± 5.42	12.21 ± 4.92	14.51 ± 4.38	14.78 ± 4.78	
	Range	[1; 23]	[1; 23]	[3; 23]	[1; 24]	
June cCTt Score	M±SD	15.66 ± 4.26	14.91 ± 4.75	16.12 ± 3.92	16.87 ± 4.51	
	Range	[3; 23]	[3; 24]	[2; 23]	[4; 24]	

Please note that the grade 3 students in the "control" group (i.e. those who have not had access to CS education between the pre and the post-tests) have slightly higher scores than those in the CS-education classes but this difference is not significant according to an ANOVA neither in the pre-test ($\Delta = 0.158pts$, $p = 0.8502$, Cohen's $D = 0.031$), nor in the post-test ($F(1) = 0.77$, $p = 0.38$, $\Delta = 0.752pts$, Cohen's $D = 0.167$)

A.2: Hierarchical linear regression model for study 1b

A.3: Analysis of variance on the student learning data for study 1a

Table 11 Hierarchical linear model for student learning (dependent variable: Delta between pre- and post- test scores, n=989) with significant variables in bold

	Estimate	95% ci	Std. error	Degrees of freedom	t-value	p-value
(Intercept)	**7.11**	[5.08, 9.15]	1.04	922	6.86	**$p < 0.0001$**
Pre-test score	**− 0.379**	[−0.52, −0.24]	0.0722	922	−5.25	**$p < 0.0001$**
Gender (girls)	0.697	[−1.62, 3.01]	1.18	922	0.591	0.555
Grade (4)	1.15	[−1.97, 4.27]	1.55	45	0.741	0.462
NCS	0.122	[−0.43, 0.68]	0.275	45	0.442	0.661
Pre-test score:gender (girls)	−0.0221	[−0.20, 0.16]	0.0906	922	−0.243	0.808
Pre-test score:grade (4)	−0.0383	[−0.24, 0.16]	0.101	922	−0.377	0.706
Gender (girls):grade (4)	−0.880	[−4.52, 2.76]	1.86	922	−0.474	0.636
Pre-test score:NCS	−0.00386	[−0.04, 0.04]	0.0198	922	−0.194	0.846
Gender (girls):NCS	−0.346	[−0.96, 0.26]	0.311	922	−1.11	0.267
Grade 4:NCS	−0.260	[−1.22, 0.70]	0.478	45	−0.544	0.589
Pre-test score:Gender (girls):Grade (4)	0.0308	[−0.23, 0.29]	0.131	922	0.235	0.814
Pre-test score:gender (girls):NCS	0.0224	[−0.03, 0.07]	0.0255	922	0.876	0.381
Pre-test score:grade (4):NCS	0.00979	[−0.05, 0.07]	0.0326	922	0.300	0.764
Gender (girls):grade (4):NCS	0.195	[−0.91, 1.30]	0.562	922	0.347	0.729
Pre-test score:gender (girls):grade (4):NCS	−0.0129	[−0.09, 0.07]	0.0412	922	−0.313	0.755

$R^2 = 0.285$, RMSE $= 2.89$, AIC $= 5132$, BIC $= 5225$, log-likelihood$= -2547$. *NCS* number of CS activities taught. Random effects $\sigma^2 = 8.85$, $\tau_{class} = 0.00$, $\tau_{school} = 1.50$ for 55 classes in 7 schools

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 33 of 43

Appendix B: Appendix for study 2

B.1: Structural equation model for the effect of student-related variables on their perception of the discipline (study 2)

Table 12 ANOVA of student learning data with Benjamini–Hochberg p-value correction and minimum effect size (Cohen's D) that can be detected with the sample

Independent variable	Sum of squares	Degrees of freedom	F	Number of groups	Residual degrees of freedom	p	Min Cohen's D	Significant difference (effect size and Dunn's post hoc test for interaction effects)
Time	3357	1	137.9	2	2636	0.0000	0.109	Post-test > Pre-test $\Delta = +2.256pts$, $p < 0.0001$, Cohen's $D = 0.457$
Grade	4007	1	166.3	2	2636	0.0000	0.109	Grade 4>3 $\Delta = 2.468pts$, Cohen's $D = 0.502$
Gender	199	1	7.8	2	2636	0.0052	0.109	Boys > Girls $\Delta = 0.551pts$, $p = 0.0015$, Cohen's $D = 0.109$
Time:grade	7396	3	108.0	4	2634	0.0000	0.129	See Fig. 3
Time:gender	3565	3	48.9	4	2634	0.0000	0.129	Pre-test Boys > Girls $\Delta = 0.664pts$, $p = 0.0079$, Cohen's $D = 0.131$ Post-test Boys ~ Girls $\Delta = 0.438pts$, $p = 0.0744$, Cohen's $D = 0.091$
Grade:gender	4248	3	58.9	4	2634	0.0000	0.129	Grade 3 Boys > Girls $\Delta = 0.725pts$, $p = 0.004$, Cohen's $D = 0.145$ Grade 4 Boys ~ Girls $\Delta = 0.469pts$, $p = 0.0604$, Cohen's $D = 0.098$
Time:grade:gender	7660	7	48.1	8	2630	0.0000	0.148	See Fig. 3

El-Hamamsy *et al. International Journal of STEM Education (2023) 10:60*

Page 34 of 43

B.2: Structural equation model for the link between perception and learning (study 2)

Table 13 Unstandardised factor loadings and regression slopes for the perception SEM (*n* = 2116, November 2021)

		Estimate	Std. Err.	Z	p	R^2
Factor loadings						
Discipline perception	Tablets	0.83	0.11	7.37	0.000	0.41
	Robots	0.76	0.10	7.85	0.000	0.37
	CS	1.98	0.70	2.82	0.005	0.81
CS	CS interest	0.17	0.05	3.25	0.001	0.25
	CS utility	0.16	0.05	3.19	0.001	0.25
	CS self-efficacy	0.20	0.06	3.31	0.001	0.28
Role models	Role models—teacher	0.17	0.01	15.60	0.000	0.13
	Role models—mother	0.23	0.01	25.37	0.000	0.28
	Role models—father	0.30	0.01	32.03	0.000	0.36
	Role models—student	0.13	0.01	15.47	0.000	0.11
	Role models—nobody	−0.24	0.01	−22.85	0.000	0.41
Tablets	Tablets interest	0.25	0.03	7.97	0.000	0.22
	Tablets utility	0.30	0.03	9.29	0.000	0.37
	Tablets self-efficacy	0.24	0.03	8.96	0.000	0.31
Robots	Robots interest	0.44	0.04	10.70	0.000	0.33
	Robots utility	0.36	0.04	9.36	0.000	0.29
	Robots self-efficacy	0.49	0.04	11.34	0.000	0.42
Regression slopes						
Discipline perception	Role models	0.15	0.05	2.94	0.003	
CS	Role models	0.30	0.12	2.41	0.016	
	Number of CS education periods	−0.01	0.01	−0.62	0.536	
Role models	Number of CS education periods	−0.01	0.01	−1.18	0.238	
	Number of ICT instruction periods	−0.00	0.00	−0.13	0.896	
	Number of robotics instruction periods	0.03	0.02	1.22	0.224	
Tablets	Number of ICT instruction Periods	−0.00	0.00	−0.80	0.426	
Robots	Number of robotics instruction periods	0.00	0.01	0.06	0.949	
CS interest	Gender (boys = 0 girls = 1)	−0.20	0.03	−5.75	0.000	
	Grade	−0.02	0.02	−1.12	0.262	
	General self-efficacy	0.09	0.03	3.29	0.001	
CS utility	Gender (boys = 0 girls = 1)	−0.05	0.03	−1.60	0.110	
	Grade	0.09	0.02	5.49	0.000	
	General self-efficacy	0.09	0.02	3.99	0.000	
CS self-efficacy	Gender (boys = 0 girls = 1)	−0.01	0.04	−0.33	0.741	
	Grade	0.01	0.02	0.74	0.459	
	General self-efficacy	0.15	0.03	5.19	0.000	
Role nodels—teacher	Gender (boys = 0 girls = 1)	0.06	0.02	2.98	0.003	
	Grade	−0.04	0.01	−3.78	0.000	
Role models—mother	Gender (boys = 0 girls = 1)	0.01	0.02	0.35	0.723	
	Grade	−0.06	0.01	−5.84	0.000	
Role models—father	Gender (boys = 0 girls = 1)	−0.06	0.02	−2.84	0.005	
	Grade	−0.01	0.01	−1.07	0.286	
Role models—student	Gender (boys = 0 girls = 1)	0.00	0.02	0.04	0.971	
	Grade	−0.00	0.01	−0.56	0.572	
Role models—nobody	Gender (boys = 0 girls = 1)	0.01	0.02	0.43	0.667	
	Grade	−0.03	0.01	−3.70	0.000	

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 35 of 43

Table 13 (continued)

		Estimate	Std. Err.	Z	p	R^2
Tablets interest	Gender (boys = 0 girls = 1)	−0.09	0.03	−2.70	0.007	
	Grade	−0.04	0.02	−2.45	0.014	
	General self-efficacy	−0.00	0.02	−0.07	0.946	
Tablets utility	Gender (boys = 0 girls = 1)	−0.05	0.03	−1.66	0.097	
	Grade	0.03	0.01	2.33	0.020	
	General self-efficacy	0.08	0.02	3.91	0.000	
Tablets self-efficacy	Gender (boys = 0 girls = 1)	−0.07	0.03	−2.55	0.011	
	Grade	−0.01	0.01	−1.07	0.283	
	General self-efficacy	0.06	0.02	3.29	0.001	
Robots interest	Gender (boys = 0 girls = 1)	−0.32	0.04	−7.32	0.000	
	Grade	0.03	0.02	1.44	0.151	
	General self-efficacy	0.09	0.03	2.50	0.012	
Robots utility	Gender (boys = 0 girls = 1)	−0.19	0.04	−4.95	0.000	
	Grade	0.07	0.02	3.78	0.000	
	General self-efficacy	0.16	0.03	4.98	0.000	
Robots self-efficacy	Gender (boys = 0 girls = 1)	−0.25	0.04	−5.90	0.000	
	Grade	0.02	0.02	0.71	0.477	
	General self-efficacy	0.19	0.03	5.52	0.000	

Appendix C: Appendix for study 3

C.1 Structural equation model between students in schools with and without access to CS education (study 3)

Table 14 Unstandardised regression parameters for the perception and background to learning SEM ($n = 1583$, November 2021, $\chi^2(124) = 221.462$, $p < 0.001$, chi^2/df = 1.79, CFI = 0.951, TLI = 0.923, RMSEA = 0.022, ci = [0.017, 0.027], SRMR = 0.026)

		Estimate	Std. Err.	Z	p	R^2
Factor loadings						
Discipline perception	Tablets	0.76	0.10	7.70	0.000	0.37
	Robots	0.82	0.11	7.22	0.000	0.41
	CS	1.51	0.38	3.93	0.000	0.71
CS	CS interest	0.21	0.05	4.62	0.000	0.28
	CS self-efficacy	0.27	0.06	4.57	0.000	0.32
Role model	Role models—teacher	0.18	0.01	14.77	0.000	0.17
	Role models—mother	0.24	0.01	23.17	0.000	0.30
	Role models—father	0.30	0.01	29.01	0.000	0.37
	Role models—student	0.14	0.01	13.72	0.000	0.11
	Role models—nobody	−0.23	0.01	−19.99	0.000	0.39
Tablets	Tablets interest	0.27	0.04	7.73	.000	0.24
	Tablets utility	0.36	0.03	10.34	0.000	0.44
	Tablets self-efficacy	0.27	0.03	9.00	0.000	0.36
Robots	Robots interest	0.42	0.05	8.92	0.000	0.34
	Robots utility	0.36	0.04	8.38	0.000	0.31
	Robots self-efficacy	0.46	0.05	9.50	0.000	0.41
Regression slopes						
Discipline perception	Role models	0.17	0.06	2.94	0.003	
CS	Role Models	0.13	0.10	1.33	0.182	
	Number of CS education periods	−0.01	0.01	−1.09	0.275	

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 36 of 43

Table 14 (continued)

		Estimate	Std. Err.	Z	p	R²
CS interest	Gender (boys = 0, girls = 1)	−0.19	0.04	−4.85	0.000	
	Grade	−0.02	0.02	−1.05	0.292	
	General self-efficacy	0.09	0.03	3.08	0.002	
CS utility	Gender (boys = 0, girls = 1)	−0.06	0.04	−1.69	0.091	
	Grade	0.05	0.02	2.71	0.007	
	General self-efficacy	0.10	0.03	3.76	0.000	
CS self-efficacy	Gender (boys = 0, girls = 1)	0.05	0.05	1.10	0.270	
	Grade	−0.01	0.02	−0.63	0.527	
	General self-efficacy	0.14	0.03	4.15	0.000	
Role models—teacher	Gender (boys = 0, girls = 1)	0.08	0.02	3.30	0.001	
	Grade	−0.06	0.01	−4.79	0.000	
Role models—mother	Gender (boys = 0, girls = 1)	−0.00	0.02	−0.13	0.900	
	Grade	−0.06	0.01	−5.55	0.000	
Role Models - father	Gender (boys = 0, girls = 1)	−0.07	0.03	−2.56	0.010	
	Grade	−0.01	0.01	−1.16	0.247	
Role models—student	Gender (boys = 0, girls = 1)	0.01	0.02	0.55	0.585	
	Grade	−0.00	0.01	−0.23	0.821	
Role models— nobody	Gender (boys = 0, girls = 1)	0.01	0.02	0.68	0.497	
	Grade	−0.02	0.01	−2.36	0.018	
Tablets interest	Gender (boys = 0, girls = 1)	−0.07	0.04	−1.86	0.063	
	Grade	−0.04	0.02	−2.08	0.037	
	General self-efficacy	0.01	0.03	0.37	0.713	
Tablets utility	Gender (boys = 0, girls = 1)	−0.05	0.04	−1.45	0.148	
	Grade	0.02	0.02	1.10	0.270	
	General self-efficacy	0.07	0.02	2.97	0.003	
Tablets self-efficacy	Gender (boys = 0, girls = 1)	−0.06	0.03	−2.17	0.030	
	Grade	−0.01	0.02	−0.63	0.527	
	General self-efficacy	0.08	0.02	3.58	0.000	
Robots interest	Gender (boys = 0, girls = 1)	−0.35	0.05	−6.83	0.000	
	Grade	0.04	0.03	1.35	0.176	
	General self-efficacy	0.09	0.04	2.15	0.031	
Robots utility	Gender (boys = 0, girls = 1)	−0.16	0.04	−3.67	0.000	
	Grade	0.06	0.02	2.44	0.015	
	General self-efficacy	0.16	0.04	4.20	0.000	
Robots self-efficacy	Gender (boys = 0, girls = 1)	−0.21	0.05	−4.39	0.000	
	Grade	0.00	0.03	0.04	0.970	
	General self-efficacy	0.20	0.04	5.31	0.000	
Role models	Number of CS education periods	0.00	0.00	0.13	0.896	
Tablets	Number of ICT education periods	−0.00	0.00	−0.52	0.601	
Robots	Number of robotics education periods	−0.01	0.02	−0.79	0.429	
Percentage (/100)						0.136
	CS perception	−0.29	1.31	−0.22	0.824	0.707
	Tablets perception	−0.04	0.99	−0.04	0.969	0.372
	Robots perception	0.96	1.11	0.86	0.387	0.411
	General self-efficacy	**1.54**	0.69	2.22	**0.027**	
	Gender (0 = boys, 1 = girls)	−1.57	1.06	−1.47	0.141	
	Grade	**7.53**	0.53	14.19	**0.000**	
	Number of CS education periods SI	0.09	0.13	0.70	0.483	
	Number of ICT education periods	−0.02	0.08	−0.20	0.842	
	Number of robotics education periods	−0.29	0.42	−0.68	0.496	

As CS utility did not correlate highly with interest and self-efficacy it was removed from this model

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 37 of 43

Table 15 Unstandardised factor loadings, regression slopes and intercepts for the perception SEM (*n* = 1640, June 2022)

	CS education					No CS education				
	Est.	Std. Err.	Z	p	R²	Est.	Std. Err.	Z	p	R²
Factor loadings										
CS										
CS interest	0.43	0.04	10.38	0.000	0.41	0.43	0.04	10.38	0.000	0.34
CS utility	0.35	0.04	9.04	0.000	0.24	0.35	0.04	9.04	0.000	0.24
CS self-efficacy	0.36	0.04	8.96	0.000	0.19	0.36	0.04	8.96	0.000	0.23
Tablets										
Tablets interest	0.39	0.04	10.62	0.000	0.38	0.39	0.04	10.62	0.000	0.35
Tablets utility	0.37	0.03	11.48	0.000	0.32	0.37	0.03	11.48	0.000	0.28
Tablets self-efficacy	0.31	0.03	9.25	0.000	0.28	0.31	0.03	9.25	0.000	0.39
Robots										
Robots interest	0.55	0.05	11.24	0.000	0.38	0.55	0.05	11.24	0.000	0.35
Robots utility	0.48	0.04	10.79	0.000	0.34	0.48	0.04	10.79	0.000	0.29
Robots self-efficacy	0.59	0.05	11.43	0.000	0.39	0.59	0.05	11.43	0.000	0.42
Role model										
Role models—teacher	0.19	0.02	12.67	0.000	0.16	0.19	0.02	12.67	0.000	0.14
Role models—mother	0.24	0.01	16.56	0.000	0.26	0.24	0.01	16.56	0.000	0.24
Role models—father	0.23	0.02	14.73	0.000	0.23	0.23	0.02	14.73	0.000	0.21
Role models—student	0.16	0.01	12.38	0.000	0.13	0.16	0.01	12.38	0.000	0.11
Role models—other	0.16	0.01	11.12	0.000	0.15	0.16	0.01	11.12	0	0.10
Role models—nobody	−0.21	0.01	−13.91	0.000	0.43	−0.21	0.01	−13.91	0.000	0.40
Regression slopes										
CS interest										
Gender (boys=0, girls=1)	−0.24	0.05	−5.02	0.000		−0.11	0.05	−2.14	.032	
Grade	0.02	0.02	0.82	0.411		−0.00	0.03	−0.04	.971	
General self-efficacy	0.02	0.04	0.55	0.583		0.10	0.05	2.09	.036	
CS utility										
Gender (boys = 0, girls = 1)	−0.03	0.05	−0.61	0.542		−0.10	0.05	−1.75	.080	
Grade	0.10	0.03	3.77	0.000		−0.00	0.03	−0.03	.975	
General self-efficacy	0.07	0.04	1.69	0.092		0.14	0.04	3.45	.001	
CS self-efficacy										
Gender (boys = 0, girls = 1)	−0.13	0.06	−2.15	0.031		−0.08	0.06	−1.37	.172	
Grade	0.02	0.03	0.54	0.592		0.01	0.03	0.46	.643	
General self-efficacy	0.16	0.05	3.26	0.001		0.20	0.05	3.80	.000	
Tablets interest										
Gender (boys = 0, girls = 1)	−0.13	0.04	−2.93	0.003		0.09	0.05	1.83	.067	
Grade	−0.03	0.02	−1.55	0.121		−0.02	0.02	−0.90	.367	
General self-efficacy	0.04	0.04	0.87	0.386		0.01	0.05	0.22	.824	

Table 15 (continued)

		CS education					No CS education				
		Est.	Std. Err.	Z	p	R^2	Est.	Std. Err.	Z	p	R^2
Tablets utility	Gender (boys = 0, girls = 1)	−0.03	0.05	−0.75	0.453		−0.03	0.06	−0.51	.611	
	Grade	−0.01	0.02	−0.63	0.529		−0.04	0.03	−1.42	.157	
	General self-efficacy	0.05	0.03	1.52	0.130		0.07	0.04	1.66	.098	
Tablets self-efficacy	Gender (boys=0, girls=1)	−0.18	0.04	−4.14	0.000		−0.07	0.04	−1.71	.087	
	Grade	0.02	0.02	0.88	0.377		0.00	0.02	0.18	.858	
	General self-efficacy	0.08	0.03	2.41	0.016		0.08	0.04	1.93	.054	
Robots interest	Gender (boys = 0, girls = 1)	−0.33	0.07	−4.89	0.000		−0.45	0.08	−5.72	.000	
	Grade	−0.03	0.03	−1.18	0.238		−0.06	0.04	−1.59	.111	
	General self-efficacy	0.02	0.04	0.43	0.667		0.09	0.06	1.53	.127	
Robots utility	Gender (boys = 0, girls = 1)	−0.12	0.06	−2.06	0.040		−0.03	0.07	−0.46	.643	
	Grade	−0.03	0.03	−1.25	0.212		−0.09	0.04	−2.61	.009	
	General self-efficacy	0.01	0.04	0.37	0.712		0.11	0.05	2.01	.044	
Robots self-efficacy	Gender (boys=0, girls=1)	−0.27	0.07	−3.94	0.000		−0.41	0.07	−5.72	.000	
	Grade	−0.05	0.03	−1.81	0.070		−0.01	0.04	−0.31	.753	
	General self-efficacy	0.11	0.05	2.14	0.033		0.14	0.06	2.42	.016	
Role models—teacher	Gender (boys = 0, girls = 1)	0.10	0.04	2.57	0.010		0.11	0.04	3.16	.002	
	Grade	−0.01	0.02	−0.64	0.522		0.00	0.02	0.27	.789	
Role Models—mother	Gender (boys = 0, girls = 1)	0.03	0.04	0.74	0.458		0.01	0.04	0.21	.833	
	Grade	−0.06	0.02	−3.73	0.000		−0.07	0.02	−4.62	.000	
Role models—father	Gender (boys=0, girls=1)	−0.07	0.04	−1.82	0.069		−0.01	0.03	−0.19	.849	
	Grade	−0.04	0.02	−2.32	0.020		−0.05	0.02	−2.98	.003	
Role models—student	Gender (boys = 0, girls = 1)	−0.05	0.03	−1.38	0.169		−0.07	0.04	−1.87	.062	
	Grade	0.03	0.01	2.13	0.033		−0.04	0.02	−2.74	.006	
Role models—other	Gender (boys = 0, girls = 1)	0.05	0.04	1.43	0.154		−0.01	0.03	−0.19	.848	
	Grade	0.09	0.01	5.98	0.000		0.00	0.02	0.09	.929	

Table 15 (continued)

		CS education					No CS education				
		Est.	Std. Err.	Z	p	R²	Est.	Std. Err.	Z	p	R²
Role models—nobody	Gender (boys = 0, girls = 1)	−0.02	0.02	−1.00	0.318		0.00	0.02	0.04	.969	
	Grade	−0.01	0.01	−0.73	0.466		0.03	0.01	3.10	.002	
Intercepts											
	CS interest	1.59	0.14	11.15	0.000		1.53	0.14	10.81	.000	
	CS utility	1.02	0.16	6.27	0.000		1.45	0.13	10.87	.000	
	CS self-efficacy	1.19	0.17	7.12	0.000		1.29	0.16	7.93	.000	
	Tablets interest	1.88	0.13	14.00	0.000		1.74	0.13	13.28	.000	
	Tablets utility	1.64	0.14	11.98	0.000		1.67	0.13	13.03	.000	
	Tablets self-efficacy	1.62	0.11	14.19	0.000		1.69	0.11	14.81	.000	
	Robots interest	1.71	0.15	11.19	0.000		1.45	0.16	9.14	.000	
	Robots utility	1.66	0.14	11.81	0.000		1.57	0.17	9.42	.000	
	Robots self-efficacy	1.51	0.16	9.27	0.000		1.19	0.16	7.59	.000	
	Role models—teacher	0.46	0.08	6.10	0.000		0.36	0.08	4.66	.000	
	Role models—mother	0.71	0.08	9.46	0.000		0.81	0.07	11.28	.000	
	Role models—father	0.77	0.07	10.50	0.000		0.78	0.07	11.26	.000	
	Role models—student	0.19	0.07	2.71	0.007		0.58	0.08	7.52	.000	
	Role models—other	0.01	0.07	0.18	0.855		0.32	0.08	4.22	.000	
	Role models—nobody	0.15	0.04	3.43	0.001		0.02	0.04	0.57	.566	
	Gender (boys = 0, girls = 1)	0.49	0.02	28.17	0.000		0.49	0.02	28.09	.000	
	Grade	4.61	0.04	115.25	0.000		4.63	0.04	120.76	.000	
	General self-efficacy	1.48	0.03	59.29	0.000		1.47	0.03	50.86	.000	

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 40 of 43

Declarations

Ethics approval and consent to participate
The researchers were granted ethical approval to conduct the study by the head of the Department of Education and by the Human Research Ethics Committee of the Anonymous university (HREC 033-2019).

Competing interests
The authors declare that the research was conducted without any competing interests.

Author details
[1]MOBOTS Group, EPFL, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. [2]LEARN, Center for Learning Sciences, EPFL, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. [3]Computer Human Interaction in Learning and Instruction (CHILI) Laboratory, EPFL, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. [4]University of Teacher Education (Haute Ecole Pédagogique) Vaud, Lausanne, Switzerland.

References

Alavi, M., Visentin, D. C., Thapa, D. K., Hunt, G. E., Watson, R., & Cleary, M. (2020). Chi-square for model fit in confirmatory factor analysis. *Journal of Advanced Nursing, 76*(9), 2209–2211.

Angot, C. (2013). *La dynamique de la motivation situationnelle*. Limoges: These de doctorat.

Apiola, M., Saqr, M., & López-Pernas, S. (2023). The evolving themes of computing education research: Trends, topic models, and emerging research. In M. Apiola, S. López-Pernas, & M. Saqr (Eds.), *Past, present and future of computing education research: A global perspective* (pp. 151–169). Cham: Springer International Publishing.

Avry, S., Emilie-Charlotte, M., El-Hamamsy, L., Caneva, C., Pulfrey, C., Dehler Zufferey, J., & Mondada, F. (2022). Monitoring the implementation of digital education by educators: a revised model.

Balanskat, A. & Engelhardt, K. (2015). Computer programming and coding Priorities, school curricula and initiatives across Europe. Technical report, European Schoolnet, (EUN Partnership AIBSL) Rue de Treves 61 1040 Brussels Belgium.

Bandura, A. (1986). Social foundations of thought and action. *Englewood Cliffs, NJ, 1986*(23-28).

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*(2), 117–148.

Barr, D., Harrison, J., & Conery, L. (2011). Computational thinking: A digital age skill for everyone. *Learning & Leading with Technology, 38*(6), 20–23.

Bers, M. U., Flannery, L., Kazakoff, E. R., & Sullivan, A. (2014). Computational thinking and tinkering: Exploration of an early childhood robotics curriculum. *Computers & Education, 72*, 145–157.

Bers, M. U., Strawhacker, A., & Sullivan, A. (2022b). The state of the field of computational thinking in early childhood education. OECD Education Working Papers 274.

Bers, M. U., Govind, M., & Relkin, E. (2022a). Coding as another language: Computational thinking, robotics and literacy in first and second grade. In *Computational Thinking in PreK-5: Empirical Evidence for Integration and Future Directions*, (pp. 30–38).

Beyer, S. (2014). Why are women underrepresented in Computer Science? Gender differences in stereotypes, self-efficacy, values, and interests and predictors of future CS course-taking and grades. *Computer Science Education, 24*(2–3), 153–192.

Blotnicky, K. A., Franz-Odendaal, T., French, F., & Joy, P. (2018). A study of the correlation between STEM career knowledge, mathematics self-efficacy, career interests, and career activities on the likelihood of pursuing a STEM career among middle school students. *IJ STEM Ed, 5*(1), 22.

Bocconi, S., Chioccariello, A., Kampylis, P., Dagien, V., Wastiau, P., Engelhardt, K., Earp, J., Horvath, M. A., Jasut, E., Malagoli, C., Masiulionyt-Dagien, V., & Stupurien, G. (2022). Reviewing computational thinking in compulsory education. JRC Research Reports JRC128347, Joint Research Centre (Seville site).

Brennan, K. & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. In *Proceedings of the 2012 annual meeting of the American educational research association, Vancouver, Canada*, (p 25).

Burroughs, N., Gardner, J., Lee, Y., Guo, S., Touitou, I., Jansen, K., & Schmidt, W. (2019). A Review of the literature on teacher effectiveness and student outcomes. In N. Burroughs, J. Gardner, Y. Lee, S. Guo, I. Touitou, K. Jansen, & W. Schmidt (Eds.), *Teaching for excellence and equity: Analyzing teacher characteristics, behaviors and student outcomes with TIMSS, IEA research for education* (pp. 7–17). Cham: Springer International Publishing.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming*. Sage.

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff Points in RMSEA test statistic in structural equation models. *Sociological Methods and Research, 36*(4), 462–494.

Cheryan, S., Plaut, V. C., Handron, C., & Hudson, L. (2013). The Stereotypical Computer Scientist: Gendered Media Representations as a Barrier to Inclusion for Women. *Sex Roles, 69*(1), 58–71.

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin, 143*(1), 1–35.

Coletta, V. P., & Steinert, J. J. (2020). Why normalized gain should continue to be used in analyzing preinstruction and postinstruction scores on concept inventories. *Physical Review Physical Education Research, 16*(1), 010108.

Croninger, R. G., Rice, J. K., Rathbun, A., & Nishio, M. (2007). Teacher qualifications and early learning: Effects of certification, degree, and experience on first-grade student achievement. *Economics of Education Review, 26*(3), 312–324.

Danaher, P. J., & Haddrell, V. (1996). A comparison of question scales used for measuring customer satisfaction. *International Journal of Service Industry Management, 7*(4), 4–26.

Deci, E. L., Connell, J. P., & Ryan, R. M. (1989). Self-determination in a work organization. *Journal of Applied Psychology, 74*(4), 580–590.

Denning, P. J., & Tedre, M. (2021). Computational thinking: A disciplinary perspective. *Informatics in Education, 20*(3), 361–390.

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology, 61*, 101859.

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 41 of 43

El-Hamamsy, L., Bruno, B., Avry, S., Chessel-Lazzarotto, F., Zufferey, J. D., & Mondada, F. (2022). *The tacs model: Understanding primary school teachers' adoption of computer science pedagogical content*. Educ: ACM Trans. Comput.

El-Hamamsy, L., Bruno, B., Chessel-Lazzarotto, F., Chevalier, M., Roy, D., Zufferey, J. D., & Mondada, F. (2021). The symbiotic relationship between educational robotics and computer science in formal education. *Education and Information Technologies, 26*(5), 5077–5107.

El-Hamamsy, L., Bruno, B., Kovacs, H., Chevalier, M., Dehler Zufferey, J., & Mondada, F. (2022b). A case for co-construction with teachers in curricular reform: Introducing computer science in primary school. In *Australasian Computing Education Conference*, ACE '22, (pp. 56–65), New York, NY, USA. Association for Computing Machinery.

El-Hamamsy, L., Chessel-Lazzarotto, F., Bruno, B., Roy, D., Cahlikova, T., Chevalier, M., Parriaux, G., Pellet, J.-P., Lanarès, J., Zufferey, J. D., & Mondada, F. (2021). A computer science and robotics integration model for primary school: Evaluation of a large-scale in-service K-4 teacher-training program. *Education and Information Technologies, 26*(3), 2445–2475.

El-Hamamsy, L., Monnier, E.-C., Avry, S., Chevalier, M., Bruno, B., Dehler Zufferey, J., & Mondada, F. (2023b). Modelling the sustainability of a primary school digital education curricular reform and professional development program. *Education and Information Technologies*.

El-Hamamsy, L., Pellet, J.-P., Roberts, M., Kovacs, H., Bruno, B., Zufferey, J. D., & Mondada, F. (2023). A research-practice partnership to introduce computer science in secondary school: Lessons from a pilot program. Educ ACM Trans. Comput

El-Hamamsy, L., Zapata-Cáceres, M., Barroso, E. M., Mondada, F., Zufferey, J. D., & Bruno, B. (2022c). The competent computational thinking test: Development and validation of an unplugged computational thinking test for upper primary school. *Journal of Educational Computing Research*, 07356331221081753.

El-Hamamsy, L., Zapata-Cáceres, M., Martín-Barroso, E., Mondada, F., Zufferey, J. D., Bruno, B., & Román-González, M. (2023d). The competent Computational Thinking test (cCTt): a valid, reliable and gender-fair test for longitudinal CT studies in grades 3-6. arXiv:2305.19526 [cs].

El-Hamamsy, L., Bruno, B., Dehler Zufferey, J., & Mondada, F. (2023a). Dataset for the evaluation of student-level outcomes of a primary school Computer Science curricular reform [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7489243.

Epskamp, S. (2022). semPlot: Path Diagrams and Visual Analysis of Various SEM Packages' Output. *R package version, 1*(1), 5.

European Union and Education A. a. C. E. A. (2019). *Digital education at school in Europe*. Brussels: Publications Office of the European Union.

Falkner, K., Sentance, S., Vivian, R., Barksdale, S., Busuttil, L., Cole, E., Liebe, C., Maiorana, F., McGill, M. M., & Quille, K. (2019). An International Comparison of K-12 Computer Science Education Intended and Enacted Curricula. In *Proceedings of the 19th Koli Calling International Conference on Computing Education Research*, Koli Calling '19, (pp 1–10), New York, NY, USA. Association for Computing Machinery.

Flichy, P. (2001). La place de l'imaginaire dans l'action technique. Le cas de l'internet. *Réseaux*, *109*(5).

Fofang, J., Weintrop, D., Walton, M., Elby, A., & Walkoe, J. (2020). Mutually Supportive Mathematics and Computational Thinking in a Fourth-Grade Classroom. volume 3, (pp. 1389–1396). International Society of the Learning Sciences.

Gagné, M., Forest, J., Gilbert, M.-H., Aubé, C., Morin, E., & Malorni, A. (2010). The motivation at work scale: Validation evidence in two languages. *Educational and Psychological Measurement, 70*(4), 628–646.

George, D. & Mallery, P. (2003). *SPSS for Windows step by step: a simple guide and reference, 11.0 update*. Allyn and Bacon, Boston, 4th ed edition.

Giang, C., Addimando, L., Botturi, L., Negrini, L., Giusti, A., & Piatti, A. (2023). *Have You Ever Seen a Robot?* Journal for STEM Educ Res: An Analysis of Children's Drawings Between Technology and Science Fiction.

Gretter, S., & Yadav, A. (2016). Computational thinking and media & information literacy: An integrated approach to teaching twenty-first century skills. *TechTrends, 60*(5), 510–516.

Grolnick, W. S., & Ryan, R. M. (1989). Parent styles associated with children's self-regulation and competence in school. *Journal of Educational Psychology, 81*(2), 143.

Gunbatar, M. S., & Karalar, H. (2018). Gender Differences in Middle School Students' Attitudes and Self-Efficacy Perceptions towards mBlock Programming. *European Journal of Educational Research, 7*(4), 925–933.

Guskey, T. R. (2000). *Evaluating professional development*. Corwin Press.

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching*, *8*(3).

Guzdial, M., & Morrison, B. (2016). Growing computer science education into a stem education discipline. *Communication of the ACM, 59*(11), 31–33.

Hall, J. A., & McCormick, K. I. (2022). "My Cars don't Drive Themselves": Preschoolers' Guided Play Experiences with Button-Operated Robots. *TechTrends, 66*(3), 510–526.

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science, 26*(4), 433–443.

Herman, J. L., Flores, A. R., & O'Neill, K. K. (2022). *How many adults identify as transgender in the United States?* The Williams Institute: Technical report.

Hickmott, D., & Prieto-Rodriguez, E. (2018). To assess or not to assess: Tensions negotiated in six years of teaching teachers about computational thinking. *Informatics in Education, 17*(2), 229–244.

Hinckle, M., Rachmatullah, A., Mott, B., Boyer, K. E., Lester, J., & Wiebe, E. (2020). The relationship of gender, experiential, and psychological factors to achievement in computer science. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education*, ITiCSE '20, (pp. 225–231), New York, NY, USA. Association for Computing Machinery.

Hinton, P. R. (Ed.). (2004). *SPSS explained*. New York: Routledge.

Hoeppner, B. B., Kelly, J. F., Urbanoski, K. A., & Slaymaker, V. (2011). Comparative utility of a single-item versus multiple-item measure of self-efficacy in predicting relapse among young adults. *Journal of Substance Abuse Treatment, 41*(3), 305–312.

Howard, J. L., Bureau, J., Guay, F., Chong, J. X. Y., & Ryan, R. M. (2021). Student motivation and associated outcomes: A meta-analysis from self-determination theory. *Perspectives on Psychological Science, 16*(6), 1300–1323.

Howard, J. L., Gagné, M., Van den Broeck, A., Guay, F., Chatzisarantis, N., Ntoumanis, N., & Pelletier, L. G. (2020). A review and empirical comparison of motivation scoring methods: An application to self-determination theory. *Motivation and Emotion, 44*(4), 534–548.

Howard, S. K., Schrum, L., Voogt, J., & Sligte, H. (2021). Designing research to inform sustainability and scalability of digital technology innovations. *Educational Technology Research and Development, 69*(4), 2309–2329.

Hsu, T.-C., Chang, S.-C., & Hung, Y.-T. (2018). How to learn and how to teach computational thinking: Suggestions based on a review of the literature. *Computers & Education, 126*, 296–310.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

Hubers, M. D. (2020). Paving the way for sustainable educational change: Reconceptualizing what it means to make educational changes that last. *Teaching and Teacher Education, 93*, 103083.

Hubwieser, P., Giannakos, M. N., Berges, M., Brinda, T., Diethelm, I., Magenheim, J., Pal, Y., Jackova, J., & Jasute, E. (2015). A global snapshot of computer science education in K-12 schools. In *Proceedings of the 2015 ITiCSE on Working Group Reports—ITICSE-WGR '15*, (pp. 65–83), Vilnius, Lithuania. ACM Press.

Hurt, T., Greenwald, E., Allan, S., Cannady, M. A., Krakowski, A., Brodsky, L., Collins, M. A., Montgomery, R., & Dorph, R. (2023). The computational thinking for science (CT-S) framework: operationalizing CT-S for K-12 science education researchers and educators. *International Journal of STEM Education, 10*(1), 1.

Jensen, K. J., Mirabelli, J. F., Kunze, A. J., Romanchek, T. E., & Cross, K. J. (2023). Undergraduate student perceptions of stress and mental health in engineering culture. *International Journal of STEM Education, 10*(1), 30.

Jiang, S., & Wong, G. K. W. (2022). Exploring age and gender differences of computational thinkers in primary school: A developmental perspective. *Journal of Computer Assisted Learning, 38*(1), 60–75.

Johnson, P., & Kite, B. (2020). semTable: Structural Equation Modeling Tables. *R package version, 1*, 8.

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 42 of 43

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). `semTools`: *Useful tools for structural equation modeling*. R package version 0.5-6.

Kanaki, K. & Kalogiannakis, M. (2022). Assessing Algorithmic Thinking Skills in Relation to Age in Early Childhood STEM Education. *Education Sciences*, 12(6):380. Number: 6 Publisher: Multidisciplinary Digital Publishing Institute.

Karpinski, Z., Biagi, F., & Di Pietro, G. (2021). Computational thinking, socioeconomic gaps, and policy implications. IEA Compass: Briefs in Education. Number 12. *International Association for the Evaluation of Educational Achievement*.

Kini, T. & Podolsky, A. (2016). Does teaching experience increase teacher effectiveness? A Review of the Research. Technical report, Learning Policy Institute. Publication Title: Learning Policy Institute ERIC Number: ED606426.

Klingner, J. K., Arguelles, M. E., Hughes, M. T., & Vaughn, S. (2001). Examining the Schoolwide "Spread" of Research-Based Practices. *Learning Disability Quarterly*, 24(4), 221–234.

Kong, S.-C., Chiu, M. M., & Lai, M. (2018). A study of primary school students' interest, collaboration attitude, and programming empowerment in computational thinking education. *Computers & Education, 127*, 178–189.

Kong, S.-C. & Lai, M. (2022a). Effects of a teacher development program on teachers' knowledge and collaborative engagement, and students' achievement in computational thinking concepts. *British Journal of Educational Technology*.

Kong, S.-C., & Lai, M. (2022). Validating a computational thinking concepts test for primary education using item response theory: An analysis of students' responses. *Computers & Education, 187*, 104562.

Kotsopoulos, D., Floyd, L., Dickson, B. A., Nelson, V., & Makosz, S. (2022). Noticing and naming computational thinking during play. *Early Childhood Education Journal, 50*(4), 699–708.

Kyriazos, T. A. (2018). Applied psychometrics: Writing-up a factor analysis construct validation study with examples. *Psychology, 9*(11), 2503–2530.

Ladd, H. F. & Sorensen, L. C. (2017). Returns to teacher experience: Student achievement and motivation in middle school. *Education Finance and Policy*, 12(2):241–279. Publisher: The MIT Press.

Li, Y., Schoenfeld, A. H., diSessa, A. A., Graesser, A. C., Benson, L. C., English, L. D., & Duschl, R. A. (2020). Computational thinking is more about thinking than computing. *Journal for STEM Educ Res, 3*(1), 1–18.

Lishinski, A., Narvaiz, S., & Rosenberg, J. M. (2022). Self-efficacy, Interest, and Belongingness—URM Students' Momentary Experiences in CS1. In *Proceedings of the 2022 ACM Conference on International Computing Education Research V.1*, (pp. 44–60), Lugano and Virtual Event Switzerland. ACM.

Lüdecke, D. (2022). *sjstats: Statistical Functions for Regression Models (Version 0.18.2)*.

Malespina, A., & Singh, C. (2023). Gender gaps in grades versus grade penalties: Why grade anomalies may be more detrimental for women aspiring for careers in biological sciences. *International Journal of STEM Education, 10*(1), 13.

Mannila, L., Dagiene, V., Demo, B., Grgurina, N., Mirolo, C., Rolandsson, L., & Settle, A. (2014). Computational Thinking in K-9 Education. In *Proc. Working Group Reports of the 2014 Conf. Innov. Technol. Comput. Sci. Educ. ITiCSE*, ITiCSE-WGR '14, (pp. 1–29), Uppsala, Sweden. ACM.

Mason, S. L., & Rich, P. J. (2019). Preparing elementary school teachers to teach computing, coding, and computational thinking. *Contemporary Issues in Technology and Teacher Education, 19*(4), 790–824.

Mason, S. L., & Rich, P. J. (2020). Development and analysis of the elementary student coding attitudes survey. *Computers & Education, 153*, 103898.

Master, A., Cheryan, S., Moscatelli, A., & Meltzoff, A. N. (2017). Programming experience promotes higher STEM motivation among first-grade girls. *Journal of Experimental Child Psychology, 160*, 92–106.

Master, A., & Meltzoff, A. N. (2020). Cultural stereotypes and sense of belonging contribute to gender gaps in STEM. *International Journal of Gender, Science and Technology, 12*(1), 152–198.

Master, A., Meltzoff, A. N., & Cheryan, S. (2021). Gender stereotypes about interests start early and cause gender disparities in computer science and engineering. *PNAS, 118*(48).

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017–1054.

Nouri, J., Zhang, L., Mannila, L., & Norén, E. (2020). Development of computational thinking, digital competence and 21st century skills when learning programming in K-9. *Education Inquiry, 11*(1), 1–17.

Olivier, E., Archambault, I., De Clercq, M., & Galand, B. (2019). Student self-efficacy, classroom engagement, and academic achievement: Comparing three theoretical frameworks. *Journal of Youth and Adolescence, 48*(2), 326–340.

Opps, Z. & Yadav, A. (2022). Who belongs in computer science? In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education*, pages 383–389, Providence RI USA. ACM.

Ottenbreit-Leftwich, A., & Yadav, A. (Eds.). (2022). *Computational Thinking in PreK-5: Empirical Evidence for Integration and Future Directions*. New York, NY, USA: ACM.

Pantic, K., Clarke-Midura, J., Poole, F., Roller, J., & Allan, V. (2018). Drawing a computer scientist: Stereotypical representations or lack of awareness? *Computer Science Education, 28*(3), 232–254.

Parker, M. C., Garcia, L., Kao, Y. S., Franklin, D., Krause, S., & Warschauer, M. (2022). A pair of aces: An analysis of isomorphic questions on an elementary computing assessment. In *Proceedings of the 2022 ACM Conference on International Computing Education Research- Volume 1*, ICER '22, (pp. 2–14), New York, NY, USA. Association for Computing Machinery.

Peel, A., Sadler, T. D., & Friedrichsen, P. (2022). Algorithmic explanations: An unplugged instructional approach to integrate science and computational thinking. *Journal of Science Education and Technology, 31*(4), 428–441.

Piatti, A., Adorni, G., El-Hamamsy, L., Negrini, L., Assaf, D., Gambardella, L., & Mondada, F. (2022). The CT-cube: A framework for the design and the assessment of computational thinking activities. *Computers in Human Behavior Reports, 5*, 100166.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer.

Pinheiro, J., Bates, D., & R Core Team (2022). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-157.

Plante, I., de la Sablonnière, R., Aronson, J. M., & Théorêt, M. (2013). Gender stereotype endorsement and achievement-related outcomes: The role of competence beliefs and task values. *Contemporary Educational Psychology, 38*(3), 225–235.

Polat, E., Hopcan, S., Kucuk, S., & Sisman, B. (2021). A comprehensive assessment of secondary school students' computational thinking skills. *British Journal of Educational Technology, 52*(5), 1965–1980.

Prudon, P. (2015). Confirmatory factor analysis as a tool in research using questionnaires: A critique,. *Comprehensive Psychology*, 4:03.CP.4.10.

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rachmatullah, A., Vandenberg, J., & Wiebe, E. (2022). Toward More Generalizable CS and CT Instruments: Examining the Interaction of Country and Gender at the Middle Grades Level. In *Proceedings of the 27th ACM Conference on on Innovation and Technology in Computer Science Education Vol. 1*, pages 179–185, Dublin Ireland. ACM.

Relkin, E., de Ruiter, L. E., & Bers, M. U. (2021). Learning to code and the acquisition of computational thinking by young children. *Computers & Education, 169*, 104222.

Revelle, W. (2022). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 2.2.5.

Risman, B. J. (2018). Gender as a Social Structure. In B. J. Risman, C. M. Froyum, & W. J. Scarborough (Eds.), *Handbook of the Sociology of Gender, Handbooks of Sociology and Social Research* (pp. 19–43). Cham: Springer International Publishing.

Roche, M. (2019). *L'acceptation d'un nouvel enseignement à l'école primaire : les professeurs des écoles face à la programmation informatique*. thesis, Nantes.

Román-González, M., Pérez-González, J.-C., & Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. *Computers in Human Behavior, 72*, 678–691.

El-Hamamsy *et al. International Journal of STEM Education* (2023) 10:60

Page 43 of 43

Rosenthal, R. (2010). Pygmalion Effect. In Weiner, I. B. & Craighead, W. E., editors, *The Corsini Encyclopedia of Psychology*, pages 1–2. Wiley, 1 edition.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36.

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology, 61*, 101860.

Schumacker, R. E. & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. psychology press.

Suessenbach, F., Schröder, E., & Winde, M. (2022). *Informatik für alle!* Policy report, Stifterverband: Informatikunterricht zur gesellschaftlichen Teilhabe und Chancengleichheit.

Sullivan, A. & Bers, M. U. (2016). Girls, boys, and bots: Gender differences in young children's performance on robotics and programming tasks. *JITE:IIP, 15*:145–165.

Sung, J., Lee, J. Y., & Chun, H. Y. (2023). Short-term effects of a classroom-based STEAM program using robotic kits on children in South Korea. *International Journal of STEM Education, 10*(1), 26.

Sun, L., Hu, L., & Zhou, D. (2022). Programming attitudes predict computational thinking: Analysis of differences in gender and programming experience. *Computers & Education, 181*, 104457.

Swaid, S. I. (2015). Bringing Computational Thinking to STEM Education. *Procedia Manufacturing, 3*, 3657–3662.

Toh, Y. (2016). Leading sustainable pedagogical reform with technology for student-centred learning: A complexity perspective. *The Journal of Educational Change, 17*(2), 145–169.

van den Akker, J. (2003). Curriculum Perspectives: An Introduction. In J. van den Akker, W. Kuiper, & U. Hameyer (Eds.), *Curriculum Landscapes and Trends* (pp. 1–10). Netherlands, Dordrecht: Springer.

Vandenberg, J., Rachmatullah, A., Lynch, C., Boyer, K. E., & Wiebe, E. (2021). Interaction effects of race and gender in elementary CS attitudes: A validation and cross-sectional study. *International Journal of Child-Computer Interaction, 29*, 100293.

Voogt, J., Fisser, P., Good, J., Mishra, P., & Yadav, A. (2015). Computational thinking in compulsory education: Towards an agenda for research and practice. *Education and Information Technologies, 20*(4), 715–728.

Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard: Harvard University Press.

Wang, J. & Hejazi Moghadam, S. (2017). Diversity Barriers in K-12 Computer Science Education: Structural and Social. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, SIGCSE '17, pages 615–620, New York, NY, USA. Association for Computing Machinery.

Wang, M.-T., Guo, J., & Degol, J. L. (2020). The role of sociocultural factors in student achievement motivation: A cross-cultural review. *Adolescent Research Review, 5*(4), 435–450.

Warner, J. R., Baker, S. N., Haynes, M., Jacobson, M., Bibriescas, N., & Yang, Y. (2022). Gender, Race, and Economic Status along the Computing Education Pipeline: Examining Disparities in Course Enrollment and Wage Earnings. In *Proceedings of the 2022 ACM Conference on International Computing Education Research V.1*, (pp. 61–72), Lugano and Virtual Event Switzerland. ACM.

Webb, M., Davis, N., Bell, T., Katz, Y. J., Reynolds, N., Chambers, D. P., & Sys?o, M. M. (2017). Computer science in K-12 school curricula of the 2lst century: Why, what and when? *Educ Inf Technol,22*(2), 445–468.

Weintrop, D. (2016). Defining computational thinking for mathematics and science classrooms. *J Sci Educ Technol*, 21.

Weintrop, D., Wise Rutstein, D., Bienkowski, M., & McGee, S. (2021). Assessing computational thinking: An overview of the field. *Computer Science Education, 31*(2), 113–116.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81.

Wing, J. M. (2006). Computational thinking. *Communication in ACM, 49*(3), 33–35.

Witherspoon, E. B., Schunn, C. D., Higashi, R. M., & Baehr, E. C. (2016). Gender, interest, and prior experience shape opportunities to learn programming in robotics competitions. *International Journal of STEM Education, 3*(1), 18.

Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research, 51*(1), 409–428.

Zhang, Y., Ng, O.-L., & Leung, S. (2023). Researching computational thinking in early childhood STE(A)M education context: A descriptive review on the state of research and future directions. *Journal for STEM Educational Research*.

Zucker, K. J. (2017). Epidemiology of gender dysphoria and transgender identity. *Sex Health, 14*(5), 404–411.