# Who participates in computer science education studies? A literature review on K-12 subjects

**Anna van der Meulen** [Corresp., 1] , **Felienne Hermans** [1] , **Efthimia Aivaloglou** [1, 2] , **Marlies Aldewereld** [1] , **Bart Heemskerk** [1] , **Marileen Smit** [1] , **Alaaeddin Swidan** [2] , **Charlotte Thepass** [1] , **Shirley de Wit** [1, 3]

[1] Leiden Institute of Advanced Computer Science, Leiden University, Leiden, Netherlands

[2] Open University of the Netherlands, Heerlen, Netherlands

[3] VHTO, Amsterdam, the Netherlands

Corresponding Author: Anna van der Meulen
Email address: a.n.van.der.meulen@liacs.leidenuniv.nl

Computer science education (CSEd) research within K-12 makes extensive use of empirical studies in which children participate. Insight in thedemographics of these children is important for the purpose of understanding the representativeness of the populations included. This literature review studies the demographics of subjects included in K-12 CSEd studies. We have manually inspected the proceedings of three of the main international CSEd conferences: SIGCSE, ITiCSE and ICER, of five years (2014-2018), and selected all papers pertaining to K-12 CSEd experiments. This led to a sample of 134 papers describing 143 studies. We manually read these papers to determine the demographic information that was reported on, investigating the following categories: age/grade, gender, race/ethnic background, location, prior computer science experience, socio-economic status (SES), and disability. Our findings show that children from the United States, boys and childrenwithout computer science experience are included most frequently. Race and SES are frequently not reported on, and for race as well as for disabilities there appears a tendency to report these categories only when they deviate from the majority. Further, for several demographic categories different criteria are used to determine them. Finally, most studies take place within schools. These insights can be valuable to correctly interpret current knowledge from K-12 CSEd research, and furthermore can be helpful in developing standards for consistent collection andreporting of demographic information in this community.

# Who participates in computer science education studies?
# A literature review on K-12 subjects

**Anna van der Meulen[1], Felienne Hermans[1], Efthimia Aivaloglou[1,2],
Marlies Aldewereld[1], Bart Heemskerk[1], Marileen Smit[1], Alaaeddin
Swidan[2], Charlotte Thepass[1], and Shirley de Wit[1,3]**

[1]Leiden Institute of Advanced Computer Science, University of Leiden, The Netherlands
[2]Open University of The Netherlands, Heerlen, The Netherlands
[3]Stichting VHTO, Amsterdam, The Netherlands

Corresponding author:
Anna van der Meulen[1]

Email address: a.n.van.der.meulen@liacs.leidenuniv.nl

## ABSTRACT

Computer science education (CSEd) research within K-12 makes extensive use of empirical studies in which children participate. Insight in the demographics of these children is important for the purpose of understanding the representativeness of the populations included. This literature review studies the demographics of subjects included in K-12 CSEd studies. We have manually inspected the proceedings of three of the main international CSEd conferences: SIGCSE, ITiCSE and ICER, of five years (2014-2018), and selected all papers pertaining to K-12 CSEd experiments. This led to a sample of 134 papers describing 143 studies. We manually read these papers to determine the demographic information that was reported on, investigating the following categories: age/grade, gender, race/ethnic background, location, prior computer science experience, socio-economic status (SES), and disability. Our findings show that children from the United States, boys and children without computer science experience are included most frequently. Race and SES are frequently not reported on, and for race as well as for disabilities there appears a tendency to report these categories only when they deviate from the majority. Further, for several demographic categories different criteria are used to determine them. Finally, most studies take place within schools. These insights can be valuable to correctly interpret current knowledge from K-12 CSEd research, and furthermore can be helpful in developing standards for consistent collection and reporting of demographic information in this community.

## INTRODUCTION

Computer Science Education (CSEd) is being increasingly taught in schools, starting as early as kindergarten. As a result, CSEd research, aimed at children under the age of 18, has also been on the rise. While several papers study teachers, programming environments, or curricula, many CSEd papers observe children as participants, using a variety of research methods from case studies to controlled experiments. These studies might have a large impact on policy, since many countries are currently in the process of implementing mandatory programming and computer science curricula (Barendsen et al., 2016; Hong et al., 2016). Further, reporting complete data on the participants and the settings of K12 CSEd activities is also required for activity comparison and replication across studies. Previous work focusing on CSEd studies and their reporting of activity components, including student demographic information, found that race and socio-economic (SES) status are rarely reported (McGill et al., 2018). We build upon this, by focusing specifically on participant demographic data, including information which has previously not been analyzed, such as the presence of students' disabilities. Although data on disabilities is of an especially sensitive nature and it might not always be possible or appropriate to obtain this information, we believe it is both relevant and important to consider as a demographic factor. An estimated 15% of the world population has a form of disability,[1] and inclusivity in CSEd can be helped forward if it is better understood how specific subpopulations of children can, when needed, be additionally supported. Further, we examine in depth the types of data that are reported or omitted, by assessing who is reported on (for instance, how many boys and girls for the category gender) and how a category is reported on (for instance, which indicators are used to represent SES).

We are interested in gaining an understanding of the representativeness of the children who are participating in K-12 CSEd studies. To that end, we have conducted a two-phase literature review of the proceedings of five years (2014-2018) of the SIGCSE, ITiCSE and ICER conferences. In Phase 1, we manually inspected the abstracts of all papers of at least 6 pages in length (953 papers) and then identified all papers that A) concerned K-12 (pre-university) subjects and B) involved subjects taking part in CSEd activities. 134 papers remained. Then in Phase 2, we read all the full papers, and gathered the reported demographic information of the participants in the categories: age/grade, gender, race/ethnicity, location, prior experience, SES and disabilities.

The aim of this literature review is to gain insight in the demographics of subjects participating in K-12 CSEd studies. These insights can contribute to first of all understanding the representativeness of the population included in these studies at the moment, thus correctly informing policy and enabling comparisons and replications. Second, our paper can help researchers identify which characteristics are important to report on when conducting CSEd studies.

## K-12 Computer Science Education

CSEd, and the broader *computational thinking* (Wing, 2006), has recently been made part of the curriculum in many countries (Barendsen et al., 2016; Hong et al., 2016). Especially *block-based programming languages* are currently being used in K-12 programming education extensively, most commonly using Scratch (Resnick et al., 2009) and Alice (Conway et al., 1994). In addition to blocks-based languages, robots (Kazakoff et al., 2013; Ludi et al., 2018; Swidan and Hermans, 2017) and textual programming languages (Hermans, 2020; Price and Barnes, 2015; Swidan and Hermans, 2019) are also frequently used. Other approaches use no computers at all, referred to as *unplugged* computing (Hermans and Aivaloglou, 2017).

A SIGCSE paper (Al-Zubidy et al., 2016) showed that empirical validation is common among SIGCSE papers, with about 70% of papers using some form of empiricism. There is a large variety in methods however: studies are performed both in the classroom (Grover et al., 2016b; Hickmott et al., 2017; Kazakoff and Bers, 2012) as well as in extracurricular settings (Maloney et al., 2008). Not all studies involving K-12 CSEd include subjects: in addition to small scale studies in classrooms, programs created by learners have also been analyzed by researchers. Scratch projects were used for example to explore the learning patterns of programmers in their first 50 projects (Yang et al., 2015). Aivaloglou and Hermans (2016) performed an analysis on 250.000 Scratch programs investigating the occurrence of code smells.

---

[1] https://www.un.org/development/desa/disabilities/resources/factsheet-on-persons-with-disabilities.html

**Research on Demographics in Computer Science**

There is some previous research aimed at understanding who participates in K-12 computing. Ericson and Guzdial, for example, in their exploration of the CS Advanced Placement test, specifically investigate women and non-white students (Ericson and Guzdial, 2014). Some papers also specifically looked at the reporting of demographics of students in computer science papers. In a meta-analysis from 2018 (McGill et al., 2018) 92 articles from SIGCSE, ICER, and ToCE between 2012 and 2016 were studied, analyzing which components or elements of activities are reported in K-12 studies, including data on the activities, the learning objectives, the instructors and the participants. Regarding participant demographics, they found that, while age and grade of participants are often reported on (98% and 74%), other factors are reported less, such as gender (64%), race (45%) and SES (13%). A similar study on CHI papers between 1982-2016 (Schlesinger et al., 2017) found that gender is reported in 63% of papers, race in 13% and class in 23%. The reporting of combined demographics is even more rare; only 2% of papers in the field of HCI report gender, race *and* class of subjects.

Our study expands these prior, related, works in two ways. First, our focus is on the content of the demographics, studying in depth *who* participates in CSEd studies, in addition to the information on whether the demographics are being reported on. Second, we performed a full manual analysis of an expanded set of 134 papers of the three main international conferences within the CSEd community from 2014 to 2018, in order to also gain insights in the way in which demographic information is presented and possible reasons why this information is omitted. This includes the factor 'disabilities', which previous reviews (McGill et al., 2018; Schlesinger et al., 2017) did not address as a demographic factor. A high prevalence of individuals with disabilities exists worldwide [2], yet this groups remain underrepresented in fields of computing and software engineering (Burgstahler and Ladner, 2006), and accessible technology remains lacking (Patel et al., 2020). Attention for these issues in CSEd research is rising, and several research lines on young learners with disabilities have started, often with focused questions and target groups (Hadwen-Bennett et al., 2018; Israel et al., 2020; Powell et al., 2004). Insights on support these learners need to be fully included in CSEd is growing, and can concern adapted or assistive technologies as well instructional approaches. Especially since the majority of children with disabilities takes place within classes in regular education (Van Mieghem et al., 2020), it is important to assess their representativeness within CSEd studies to further understand how inclusive CSEd education can be fostered. Together, these insights are valuable to understand the representativeness of the populations included in K-12 CSEd studies, to interpret previous and future findings, and to help guide reporting on future work.

Finally, one point to address concerns our choice of the three conferences. We believe that the proceedings of the SIGCSE, ITiCSE and ICER conferences reflect the global CSEd community well as they are often referred to as the "main international conferences", and consequently provide a valid starting point to explore the demographics of it's studies subjects. It is likely that the US will be over-represented in our findings, however, it can be considered part of our assessment to determine this and to reflect on the implications. A related point concerns the demographic category race or ethnicity. This category can be common to include, as the previous studies on demographics of K-12 computing participants confirm (McGill et al., 2018; Schlesinger et al., 2017), however, this strongly depends on the context and country. We will follow the terminology used in the studies we include, taking this together under the term "race/ethnic background". In addition, we will also explore whether the inclusion of the category of disabilities is related to context and country.

## SURVEY METHODOLOGY

**Setup of Literature Review**

The research question of this paper is: *What are the demographics of subjects who participate in CSEd studies in K-12?* In order to answer this question, a two-phase literature review was conducted. In Phase 1, all 953 papers that appeared in the proceedings of one of the three main international CSEd conferences: SIGCSE, ITiCSE and ICER between 2014 and 2018, were *inspected* using the procedure and criteria described below to determine whether they qualified for Phase 2. The resulting 134 papers were manually *analyzed* to collect the reported demographic information of each paper. The distribution of papers in each phase is shown in Table 1.

[insert Table 1 here]

---

[2]https://www.un.org/development/desa/disabilities/resources/factsheet-on-persons-with-disabilities.html

### Paper Inspection (Phase 1) and Paper Analysis (Phase 2)

In Phase 1 all full papers (6 pages or longer) which appeared in the proceedings of SIGCSE, ITiCSE and ICER in five years (2014-2018) were inspected. Phase 1 included 953 papers in total (561, 267 and 125 respectively for the three conferences). These papers were inspected by reading their title and abstract. The first papers were inspected together by the group of nine authors to evaluate the inspection process, after which the rest of the papers were divided among all authors. Papers were inspected individually, but uncertainties concerning whether a paper qualified for Phase 2 were noted down and discussed among the group of authors together until agreement was reached. For papers to be included in Phase 2, we set two criteria:

**K-12** Participants in the reported study should be K-12 level: in kindergarten, elementary school, middle school or high school. Papers that described studies of which some of the participants were K-12 while others were already in university were also considered.

**Computer Science Education activity** Participants in the reported study should be actively involved in CSEd activities. We excluded studies in which children did not actively engage in programming or other CSEd activities, for example, studies in which only children's attitude towards programming or computers was measured without any intervention.

In Phase 2 the papers selected during the first phase were analyzed. This concerned 134 papers (13% of all SIGCSE, ITiCSE and ICER 2014 to 2018 papers). Each paper was carefully read in full to gather the demographic information that was reported on or to note the lack of report of information. The specific demographic categories are described in the next section. The same process as in Phase 1 was followed, where the first three papers were read together by all authors to practice the process. Next, the papers were divided among the nine authors. Uncertainties were noted down and discussed by the authors in a weekly group meeting. In these meetings, demographic information that was inferred and not explicitly stated in the papers, such as information on location, was checked and agreed upon by the authors. Moreover, coding conventions were set for demographic information that was not uniformly reported on, such as age and school system levels, race/ethnicity and socio-economic status.

### Demographic Categories

The demographic categories were determined at the start of Phase 2, observing the categories commonly reported in the included papers. We confirmed the demographic categories that McGill *et al.* found (McGill et al., 2018): ages/grades, gender, race/ethnic background, location, prior experience and SES, and included the factor disabilities.

## RESULTS

This section presents the results of our analysis of the 134 papers reporting on CSEd studies using K-12 students as subjects. We found 9 papers in the sample that reported on multiple studies, and we decided to classify them separately, because they occassionally reported on different samples. As such in the remainder of the paper, we report on the number of *studies*, not the number of papers. The 134 papers in total report on 143 studies. Table 2 presents an overview of the number of studies reporting on the different demographic categories. Not all conferences were represented equally, as can be seen in Table 1. This is mainly due to the fact that SIGCSE is bigger in terms of papers than the other two conferences. Our sample represents the conferences quite equally: 16% of SIGCSE papers were included, 9% of ITiCSE papers and 14% of ICER papers.

Before we present the results of the demographic categories, we briefly refer to the context in which the studies are performed. All but two studies report on the context where the experiment was run. The majority of studies (88 or 62%) took place within a school, while 49 (34%) took place outside of school, for example in a coding club or summer camp. Four studies report performing experiments or interventions both in a school and in another context.

[insert Table 2 here]

### Age and Grade

The ages of the participants, or their grade, is reported in almost all papers in our sample: 134 studies (94%) report this information. Since the majority of papers refer to the classification of the US school

169  system of elementary, middle, and high-school, we followed this classification as well in order to obtain an
170  overview. Information from papers that reported on a different grade system or only on age was manually
171  classified. We classified Kindergarten to grade 6 as elementary school, grades 6 to 8 as middle school
172  and grades 9 to 12 as high school. Grade 6 is within the US classified as either elementary or middle
173  school, depending on the specific school. Consequently, where school was provided in a study for grade
174  6 we follow this and noted either elementary or middle school. Where school level was not provided,
175  we classified grade 6 as elementary school unless it was indicated as part of a range that includes higher
176  grades a well (for instance, participants being from grades 6-8), in which case we classified the whole
177  range as middle school. Middle school and high school participants are most common in the papers,
178  representing 37% and 31% of studies respectively. Some papers report on a variety of subjects in their
179  studies, for example one paper reported on a study with participants from pre-school to university (Martin
180  et al., 2017).

181  **Gender**
182  Gender is relatively well-reported, with 92 studies (64%) providing the gender of their participants. As
183  Figure 1 shows, 60 studies (44 + 16), which together represent 65% of studies that report on gender, have
184  over 55% male participants. Even though boys are over-represented in general, girls are over-represented
185  in studies focusing specifically on gender: out of the 12 studies with less than 25% male participants, 6
186  consisted exclusively of female students. Further, the majority (10) of the studies with less than 25% male
187  participants took place in camps or after school. In the studies with less than 25% female participants
188  this pattern was not visible. Besides the studies that included just one gender, there are also studies
189  that included gender in their research purpose (22% of the studies that reported gender). Overall, male
190  participants appear over-represented in the studies which report gender, even though there are several
191  papers with a special-focus on female participants. Finally, two studies report in addition to male and
192  female the category other, which includes 1% (Paspallis et al., 2018) and 5% (Schanzer et al., 2018).
193       [insert Figure 1 here]

194  **Race/Ethnic background**
195  With 46 studies (34%) reporting on the race or ethnic background of their participants, this category is
196  reported on in approximately half as many studies as gender. Out of these 46, 37 studies provide concrete
197  quantifiable information. The remaining studies for instance indicate race about a proportion of their
198  participants but leave a large proportion as "unknown" (Wang and Hejazi Moghadam, 2017) or refer
199  merely to different countries (Srikant and Aggarwal, 2017). The quantifiable information from the 37
200  studies is represented in Figure 2, showing that white students form the minority of participants, with
201  13 studies (28%) reporting 0 to 25% white students. Figure 3 depicts the number of races or ethnicity's
202  included in the studies that provide this information.
203       As is the case for gender observations, there are also a number of papers reporting on interventions,
204  tools or programs specifically aimed at ethnic minority students. Out of the 46 studies that reported
205  race/ethnicity, 15 (or 33%) studies included race/ethnicity in their research purpose. Some of these
206  studies that target a specific race/ethnic group also target a specific gender, focusing on African-American
207  girls (Thomas, 2018; Van Wart et al., 2014), or American-Indian boys (Searle and Kafai, 2015).
208       [insert Figure 2 here]
209       [insert Figure 3 here]

210  **Socio-economic status (SES)**
211  SES is reported in 28 studies (20% of studies). Figure 4 gives an overview of what is being reported in
212  these studies. The figure shows that there is no clear focus on either low or high SES students, and that
213  many papers explicitly stated the SES of subjects was mixed. Out of the 28 studies that report on SES, 8
214  (29%) included SES within their research purpose.
215       We also found that SES was approached differently across the studies. Various indicators were used,
216  including income or poverty rate, as well as eligibility of the sample or the population for free lunches or
217  other forms of extra funding, which all represent slightly different aspects of SES.
218       [insert Figure 4 here]

**Disabilities**

With only 6 studies (4%) reporting on disabilities of their participants, this category is the least reported on. Out of the 6, two studies named visually impaired children (Kane and Bigham, 2014; Ludi et al., 2018), and one named children with autism (Johnson, 2017). In one of the studies that included children with visual impairments (Kane and Bigham, 2014) and in the study that mentioned children with autism (Johnson, 2017), these target groups are part of the context in which the research takes place (a camp for students with autism and a camp for visually impaired children, resp.). Consequently, the information on the prevalence of these disabilities seems inferred from the research context and not explicitly acquired. In addition, no other details (such as specific types within these syndromes) are provided. The other study on children with visual impairments (Ludi et al., 2018) provides somewhat more information, indicating that the participants were recruited across the US, from various educational situations (homeschooled, public school, schools for the blinds), and furthermore that 58% had low vision and the remaining children were blind.

The other three papers, out of the 6 that reported on disabilities, remarked at a more general level that "students with various disabilities" participated in their study (Paramasivam et al., 2017) or that there were children with "special needs or learning difficulties" (Grover et al., 2016b) or from "special education" (Grover et al., 2016a). The paper by Paravisam et al. (2017) provided a wide range of disabilities included in their sample ("deafness, low vision or blindness, Cerebral Palsy, Muscular Dystrophy, Ollier's disease, Attention Deficit Disorder, Asperger's Syndrome, and other autism spectrum disorders or learning disabilities"), without however specifying which disability was present for how many students. All 6 studies that report on disabilities concern research conducted in the US.

Two additional papers reported comparable cognitive background information on their participants by describing that the children in the study were gifted (Friend et al., 2018), or were students in a school which is also a center for gifted children (Daily et al., 2014) (which is usually not seen as a disability, consequently we did not include these 2 papers in the 6 papers of this category).

**Location**

Location is a multi-faceted factor, which (McGill et al., 2018) describes as *"Specific locations, including city, state, and country"*. Country is reported on relatively often, with 96 studies (67%) explicitly stating the country in which the study is being executed. For 42 studies (29%), the location of the study can be deduced with reasonable certainty from the description of the setup in the paper, for example when the authors work at a university in the US and state they have executed the study themselves.

Combining papers that explicitly report subjects from the United States (65) and papers where we deduced the US as location from context (36), we find 101 studies (71% of all studies) in which children from the US participate. This overshadows studies conducted in Europe (19 or 13%) and Asia & Australia (7 or 5%). Location more specific than the country of the study is reported on in about one in three papers (46 studies). When it is reported, however, it is often with different terms focused on different aspects. For example, some papers state the location along with more information about the residents, such as *"two tribes of an American Indian community outside of Phoenix, Arizona"* (Kafai et al., 2014), while others are more generic and use broad locations such as *"Western Germany"* (Pasternak, 2016) or *"a large Northeast city"* (Deitrick et al., 2015).

**Prior Experience**

Prior experience with an aspect of computer science is reported in 52 studies (37%). However, it is typically reported on loosely using phrases such as *"some children had coding experience"* (Broll et al., 2017), *"students did not begin the class with strong experience in computers and computing"* (Freeman et al., 2014) or *"most had little to no experience with robotics design or programming."* (Ludi et al., 2018) which leaves room for interpretation as to what *some*, *strong* or *most* means exactly. Figure 5 shows the distribution of the studies over different categories of the experience of subjects. The category 'mixed' pertains to studies for which it was explicitly reported that some of the participants have experience while others do not. As can be seen in Figure 5, the largest category of studies (22, 42% of studies reporting on prior experience) reports on subjects without any experience. In some cases, papers reported on excluding children with prior knowledge of computer science through a pretest or self-assessment, for example (Zhi et al., 2018).

[insert Figure 5 here]

272 **CONCLUSIONS**

273 With this literature review we aimed to gain insight in the demographics of subjects participating in K-12
274 CSEd studies, in order to understand the representativeness of the populations included, and in order to
275 identify which characteristics empirical studies can report on. Our insights are discussed below.

276 **Main Overview of Reported Information and Included Populations**

277 First, the reported demographic information shows that children from the United States, boys and children
278 without experience appear over-represented. Second, we found that demographic information in general
279 and specific categories is often missing. Especially race or ethnic background, disabilities and SES
280 of participants are frequently not reported on, as previous studies also showed (McGill et al., 2018;
281 Schlesinger et al., 2017). Finally, most studies take place within schools. These insights are further
282 discussed below, together with other observations on the inclusion and report of demographic factors.

283 **US Focus**

284 As we expected, the category location showed that children from the US were over-represented in the
285 studies in SIGCSE, ITiCSE and ICER we reviewed. This limits the representativeness of the findings
286 of these studies first of all because of curriculum differences world-wide in the organization of K-12
287 education. In addition, in many European countries (including Sweden, Germany, Finland and the
288 Netherlands) there are no or very few private schools, and homeschooling is in European countries often
289 prohibited (in Greece, Germany and the Netherlands) or hardly practiced (in France). These differences
290 can make it difficult to generalize findings. A second implication of the over-representativeness of the US
291 is that the reporting is also distinctly US flavored. For example, when discussing SES, a paper stated that
292 a school has *"eligibility for Title I funding"* (Ibe et al., 2018) or that children received *"free or reduced*
293 *lunch"* (Buffum et al., 2016; Hansen et al., 2017) referring to distinct US policies that people outside the
294 US might not be familiar with. Similarly, a concept such as an 'urban' area (Grover et al., 2016b; Tsan
295 et al., 2018) does not have the same connotation in all countries; it can refer to a school in a poor inner
296 city area or a well-educated high-income neighborhood.

297 Interpretation and generalization could be strengthened if empirical studies (including those from
298 the US) give a clear report on factors of location and SES. This can be done either in way that is easy
299 to understand without an in-depth knowledge of a specific school system and context, or by providing a
300 brief description of the educational system in a country. Some of the included papers provide examples of
301 this approach, for example when describing the specific setting in France (Chiprianov and Gallon, 2016).
302 Further, it is important to be aware of the high presence of children from the US in studies in SIGCSE,
303 ITiCSE and ICER, and to interpret findings with this in mind or explicitly look for studies from other
304 locations.

305 **Inconsistent Criteria and School-level reporting**

306 We observed that, for several demographic categories, different criteria were being used by different
307 papers. This was most visible for SES, where some papers reported low SES based on eligibility for lunch
308 programs (Buffum et al., 2016) while others used poverty rates (Feaster et al., 2014) or the economic
309 and educational level of parents (Ko et al., 2018). The same is true for prior experience. Some papers
310 excluded children with too much prior knowledge (Zhi et al., 2018), some papers named experience with
311 the tool or language being used (Joentausta and Hellas, 2018) while others explicitly separated experience
312 with the tool from computer science experience in general (Smith et al., 2014). To increase the clarity
313 of future studies, it would be beneficial for our community to establish recommended guidelines for the
314 criteria to be used for SES or prior experience in K-12 studies.

315 Further, both in the case of race/ethnic background and disabilities it can be considered whether a
316 bias is occurring in reporting this information. First, in the case of race/ethnic background, the studies
317 that provide quantifiable information on this demographic together suggests that white students form
318 the minority, which is not in line with the American population (to which a large percentages of our
319 studies pertain), where about 50% of children enrolled in public elementary and secondary schools are
320 white.[3] It is possible that race or ethnic background is more often reported when participants do not
321 belong to the racial/ethnic majority, or, specific to the current research setting, when their background is
322 not congruent with expectations of computer science classroom populations. This is further confirmed

---
[3]https://nces.ed.gov/programs/raceindicators/indicator_rbb.asp

by the large number of studies that report on mixed-racial experimental groups. Here as well, a higher number of mixed-racial groups is reported compared to what can be expected from typical classrooms in the US (Stroub and Richards, 2013). Second, a similar type of reporting bias might be occurring for the factor disabilities, which are mainly being reported when the study targets or focuses on children with disabilities and as such all children in the study undoubtedly have a disability (such as in a camp for blind children). Here as well, it seems unlikely that only 6 studies reporting on disabilities is an accurate representation of the population, where worldwide approximately 15% has a disability [4]. However, for both of these factors (race/ethnic background and disabilities) the lack of consistent reporting might also be related to the especially sensitive nature of these factors, which is further discussed below.

Finally, some papers reported the demographics of the study not at the level of the study participants, but at the level of the school. For example: *"At the elementary school where we collected the data, the student body is roughly 53% African-American, 33% Caucasian, and 14% Hispanic, Latino, Native American, Asian, or mixed race. Approximately 47.4% of the students receive free or reduced cost lunch"* (Buffum et al., 2016). This presents only an indication of the exact study sample, which might differ because of factors such as selection bias, with some students or teachers being more willing or available to participate. At the same time, school level reporting is preferable over no reporting at all. Further, as discussed in the section below on ethical and legal considerations, it can be a good alternative when dealing with practical or ethical constraints.

**Disabilities as a Demographic Factor**

In this review we explored the inclusion of 'disabilities' as a demographic factor, which was not included in previous reviews on CSEd research populations (McGill et al., 2018; Schlesinger et al., 2017). Only 6 studies reported on the presence of disabilities, referring for instance in a relatively general manner to children with visual impairments or to learners with special needs. Usually how this information was obtained was not indicated, though in some cases this was inferred from the targeted research context (such as a camp for children with autism). It is likely that the low and relatively general reporting on disabilities can be explained by a lack of acknowledgment of this factor as a standard demographic (as the absence of mentioning it in the previous reviews also suggests), and, related, the sensitivity of these data (Fernandez et al., 2016). The need to increase inclusivity of CSEd, especially concerning learners with impairments (Burgstahler and Ladner, 2006), stresses the importance of considering how a balance can be found in reporting on this information in a more consistent, but also feasible, way. A starting point might be to consistently refer to the consideration of the presence of disabilities within a sample, adding information where possible at an appropriate level of detail.

Further assessment of how studies outside of the conferences of our focus handle and report on the presence of disabilities can be helpful. Although a thorough exploration falls outside of the scope of this review, some suggestions can be found in other CSEd journals or conferences within the topic of learners with visual impairments. Similar to the paper by Ludi et al (2018) included in our review, some of the often small scale studies within this topic tend to elaborate both on the specific type of impairment (including for example being blind from birth, or partially sighted) and on the recruitment of the participants (through special education schools or personal contacts) (Jašková and Kaliaková, 2014; Kabátová et al., 2012; Milne and Ladner, 2018; Morrison et al., 2018). An interesting option that is being applied is to elaborate on a participants' disability in terms of use of computer, assistive technologies, or tools (Jašková and Kaliaková, 2014; Milne and Ladner, 2018). Although such level of detail can be highly useful to interpret the findings, it is likely that, in studies taking place in larger groups and within mainstream education, obtaining this information is not always feasible or practical. In that case, more general (indicating the presence of disabilities) or school level (indicating a percentage) report can be valid alternative options, that still provide an indication of the population included. Authors could then also explain or substantiate why they report in this manner. For instance, Koushik and Kane (2019) indicate in their study on training computer science concepts that their participants have cognitive disabilities. They further describe : *"We did not collect individual diagnoses from our participants as we did not believe this personal information was relevant to our research goals"*, and provide a summary of different types of cognitive impairments included in the club of which the participants were members. Overall, a tailored approach might be needed for the report on disability, yet mentioning the factor and substantiating the information that is (not) provided can be very helpful for the field.

---

[4]https://www.un.org/development/desa/disabilities/resources/factsheet-on-persons-with-disabilities.html

### Separate Subjects Section

While examining the papers, we found that papers do not only differ in *which* demographic data were reported, but also in *how* this was presented in the paper. 53 papers (40% of 134 papers) have a dedicated section or subsection where all information on participants was placed, called "subjects", "participants", or something similar. In 35 papers (26%), information about participants was placed in larger sections, such as "experimental design", "research setup" or "methods". In the remaining 38% of the papers there was no central place in the paper for demographic information. Instead, it was placed across different sections, or was only stated in the introduction or abstract. We would advise journals and conferences that are considering to adopt guidelines for the disclosure of participant demographics to also consider suggesting a default name for the section in which this information should be placed.

### Ethical Considerations and Legal Consideration

Above the ethical issues in collecting and reporting demographic information have already been touched upon. All demographic information, but some especially (such as on ethnic background, SES, and disabilities) concern sensitive data that might not be easy to obtain due to objections from parents or schools. Further, if obtained, these data should be handled and reported carefully without the possibility of making the subjects of a study traceable. In behavioral sciences, the collection and report of demographic information of its' subjects, which also often concerns children, is standard and required. This can be helpful in identifying safe procedures for this type of data (for instance, using careful informed consent procedures and pseudonymisation of the obtained data). Furthermore, the example of behavioral sciences can also clarify the purpose of reporting demographic information that is in itself not part of the research question. The discussion on the representativeness of study samples exists in psychology as well, and one recommendation here is also to consistently report demographics in empirical studies (Rad et al., 2018).

The collection and reporting of detailed demographic data could also be hindered by legal concerns about the privacy of the study participants. This is an increasingly important issue, because interest in privacy and personal data protection has intensified, leading to the adoption of new laws, in particular in the European Union. Data sanitation techniques, including randomisation and generalisation, can be applied, aiming to prevent the re-identification of data subjects. The application of these techniques when reporting on demographic data on study participants is challenging, because the more detailed the reported data, the more identifiable the subjects become. For researchers reporting this data, it could understandably be a challenge to achieve a proper balance between on the one hand providing replicable and representative data and on the other hand protecting the privacy of the participants and complying with privacy laws. Regulations and customs from behavioral sciences could be helpful here.

### Limitations of Current Research and Directions for Future Work

This literature review focused specifically on proceedings of the SIGCSE, ITiCSE and ICER conferences. We believe these venues reflect the global CSEd community and provide the valid starting point to explore the demographics of it's studies' subjects. At the same time, there is some limitation because of this focus. Future work can look further into papers from CSEd journals, as well as focus on the specific case of local conferences. Moreover, for our analyses we assumed that each of the 143 studies has been reported on in exactly one paper in the dataset; a publication of a study in more than one papers in the three conferences would be a threat to the validity of our findings. Further, the analyses of specific demographic factors in the current research are somewhat limited, not looking for example at relations between different factors. Our goal was to provide an overview of different factors, future studies could focus on specific factors such as ethnic background to gain further insight. Finally, a threat to validity in the method of analyzing the papers should be mentioned. This was done manually by nine different researchers, without applying a method of double-coding. However, we attempted to reduce this threat by first practicing the analysis process with all researchers together and agreeing on common coding conventions, and second by regularly meeting and discussing the general experiences as well as specific doubts about the reported demographic information and especially about information that was not explicitly stated but deduced from the study context. In addition, we make the resulting dataset available to the research community for cross validating our findings.[5]

---

[5]Link to be provided through our institution's repository.

**Concluding Remarks**

With this paper, we have provided insight in the representativeness of participants in K-12 CSEd studies. Our literature review showed that certain populations (children from the US, boys, and children without previous computer science experience) are over-represented. Further, we showed that certain demographic information is not always or not consistently reported on. This is especially the case for race/ethnic background, SES, and disabilities. We are very much aware of practical, ethical, and legal challenges in collecting demographic information. However, interpretation, generalization, and replication of research findings as well as policy development could benefit greatly from, as much as possible, consistent report of characteristics of populations included. Consequently, an important recommendation would be to develop standards for the collection and report of demographic information in empirical studies in K-12 CSEd. Such guidelines, inspired by behavioral sciences, could include alternative options such as the report at school or neighborhood level, to be applied where appropriate.

## ACKNOWLEDGMENTS

## REFERENCES

Al-Zubidy, A., Carver, J. C., Heckman, S., and Sherriff, M. (2016). A (updated) review of empiricism at the sigcse technical symposium. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, SIGCSE '16, page 120–125, New York, NY, USA. Association for Computing Machinery.

Barendsen, E., Grgurina, N., and Tolboom, J. (2016). A New Informatics Curriculum for Secondary Education in The Netherlands. In *9th International Conference on Informatics in Schools: Situation, Evolution, and Perspectives*, pages 105–117. Springer.

Broll, B., Ledeczi, A., Volgyesi, P., Sallai, J., Maroti, M., Carrillo, A., Weeden-Wright, S. L., Vanags, C., Swartz, J. D., and Lu, M. (2017). A Visual Programming Environment for Learning Distributed Programming. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, pages 81–86. ACM.

Buffum, P. S., Frankosky, M. H., Boyer, K. E., Wiebe, E. N., Mott, B. W., and Lester, J. C. (2016). Empowering All Students: Closing the CS Confidence Gap with an In-School Initiative for Middle School Students. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, pages 382–387. ACM.

Burgstahler, S. and Ladner, R. (2006). An alliance to increase the participation of individuals with disabilities in computing careers. *ACM SIGACCESS Accessibility and Computing*, (85):3–9.

Chiprianov, V. and Gallon, L. (2016). Introducing Computational Thinking to K-5 in a French Context. In *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education*, ITiCSE '16, pages 112–117. ACM.

Conway, M., Pausch, R., Gossweiler, R., and Burnette, T. (1994). Alice: A Rapid Prototyping System for Building Virtual Environments. In *Conference Companion on Human Factors in Computing Systems*, CHI '94, pages 295–296. ACM.

Daily, S. B., Leonard, A. E., Jörg, S., Babu, S., and Gundersen, K. (2014). Dancing alice: Exploring embodied pedagogical strategies for learning computational thinking. In *Proceedings of the 45th ACM technical symposium on Computer science education*, pages 91–96.

Deitrick, E., Shapiro, R. B., Ahrens, M. P., Fiebrink, R., Lehrman, P. D., and Farooq, S. (2015). Using Distributed Cognition Theory to Analyze Collaborative Computer Science Learning. In *Proceedings of the Eleventh Annual International Conference on International Computing Education Research*, pages 51–60. ACM.

Ericson, B. and Guzdial, M. (2014). Measuring Demographics and Performance in Computer Science Education at a Nationwide Scale Using AP CS Data. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education*, SIGCSE '14, pages 217–222. ACM.

Feaster, Y., Ali, F., Zhai, J., and Hallstrom, J. O. (2014). Serious Toys: Three Years of Teaching Computer Science Concepts in K-12 Classrooms. In *Proceedings of the 2014 Conference on Innovation & Technology in Computer Science Education*, ITiCSE '14, pages 69–74, New York, NY, USA. ACM.

Fernandez, T., Godwin, A., Doyle, J., Verdin, D., Boone, H., Kirn, A., Benson, L., and Potvin, G. (2016). More comprehensive and inclusive approaches to demographic data collection. In *Proceedings of the ASEE 2016 Annual Conference*.

Freeman, J., Magerko, B., McKlin, T., Reilly, M., Permar, J., Summers, C., and Fruchter, E. (2014). Engaging Underrepresented Groups in High School Introductory Computing Through Computational Remixing with EarSketch. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education*, pages 85–90. ACM.

Friend, M., Matthews, M., Winter, V., Love, B., Moisset, D., and Goodwin, I. (2018). Bricklayer: Elementary students learn math through programming and art. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, pages 628–633.

Grover, S., Pea, R., and Cooper, S. (2016a). Factors influencing computer science learning in middle school. In *Proceedings of the 47th ACM technical symposium on computing science education*, pages 552–557.

Grover, S., Rutstein, D., and Snow, E. (2016b). "What Is A Computer": What Do Secondary School Students Think? In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, SIGCSE '16, pages 564–569. ACM.

Hadwen-Bennett, A., Sentance, S., and Morrison, C. (2018). Making Programming Accessible to Learners with Visual Impairments: A Literature Review. *International Journal of Computer Science Education in Schools*, 2(2):3.

Hansen, A. K., Dwyer, H. A., Iveland, A., Talesfore, M., Wright, L., Harlow, D. B., and Franklin, D. (2017). Assessing Children's Understanding of the Work of Computer Scientists: The Draw-a-Computer-Scientist Test. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, pages 279–284. ACM.

Hermans, F. (2020). Hedy: A gradual language for programming education. In *Proceedings of the 2020 ACM Conference on International Computing Education Research*, pages 259–270.

Hermans, F. and Aivaloglou, E. (2017). To Scratch or not to Scratch?: A controlled experiment comparing plugged first and unplugged first programming lessons. In *Proceedings of the 12th Workshop on Primary and Secondary Computing Education*, pages 49–56. ACM.

Hickmott, D., Prieto-Rodriguez, E., and Holmes, K. (2017). A Scoping Review of Studies on Computational Thinking in K-12 Mathematics Classrooms. *Digital Experiences in Mathematics Education*, pages 1–22.

Hong, H., Wang, J., and Moghadam, S. H. (2016). K-12 Computer Science Education Across the U.S. In *9th International Conference on Informatics in Schools: Situation, Evolution, and Perspectives*, pages 142–154. Springer International Publishing.

Ibe, N. A., Howsmon, R., Penney, L., Granor, N., DeLyser, L. A., and Wang, K. (2018). Reflections of a Diversity, Equity, and Inclusion Working Group Based on Data from a National CS Education Program. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, SIGCSE '18, pages 711–716. ACM.

Israel, M., Chung, M. Y., Wherfel, Q. M., and Shehab, S. (2020). A descriptive analysis of academic engagement and collaboration of students with autism during elementary computer science. *Computer Science Education*, 30(4):444–468.

Jašková, L. and Kaliaková, M. (2014). Programming Microworlds for Visually Impaired Pupils. *Proceedings of the 3rd international constructionism conference*, pages 1–10.

Joentausta, J. and Hellas, A. (2018). Subgoal Labeled Worked Examples in K-3 Education. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, SIGCSE '18, pages 616–621. ACM.

Johnson, C. (2017). Toward computational making with madeup. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, pages 297–302.

Kabátová, M., Jašková, , Lecký, P., and Laššáková, V. (2012). Robotic activities for visually impaired secondary school children. *3rd International Workshop, Teaching Robotics, Teaching with Robotics*, (January):22–31.

Kafai, Y., Searle, K., Martinez, C., and Brayboy, B. (2014). Ethnocomputing with Electronic Textiles: Culturally Responsive Open Design to Broaden Participation in Computing in American Indian Youth and Communities. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education*, SIGCSE '14, pages 241–246. ACM.

Kane, S. K. and Bigham, J. P. (2014). Tracking @Stemxcomet: Teaching Programming to Blind Students via 3d Printing, Crisis Management, and Twitter. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education*, SIGCSE '14, pages 247–252. ACM.

Kazakoff, E. R. and Bers, M. U. (2012). Programming a robotics context in the kindergarten classroom: The impact on sequencing skills. *Journal of Educational Multimedia and Hypermedia*, 21(4):371–391.

Kazakoff, E. R., Sullivan, A., and Bers, M. U. (2013). The effect of a classroom-based intensive robotics and programming workshop on sequencing ability in early childhood. *Early Childhood Education Journal*, 41(4):245–255.

Ko, A., Hwa, L., Davis, K., and Yip, J. C. (2018). Informal Mentoring of Adolescents About Computing: Relationships, Roles, Qualities, and Impact. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, pages 598–603.

Ludi, S., Bernstein, D., and Mutch-Jones, K. (2018). Enhanced Robotics!: Improving Building and Programming Learning Experiences for Students with Visual Impairments. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, SIGCSE '18, pages 372–377. ACM.

Maloney, J. H., Peppler, K., Kafai, Y., Resnick, M., and Rusk, N. (2008). Programming by Choice: Urban Youth Learning Programming with Scratch. In *Proceedings of the 39th SIGCSE Technical Symposium on Computer Science Education*, pages 367–371. ACM.

Martin, C., Hughes, J., and Richards, J. (2017). Learning Dimensions: Lessons from Field Studies. In *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education*, ITiCSE '17, pages 299–304. ACM.

McGill, M. M., Decker, A., and Abbott, Z. (2018). Improving Research and Experience Reports of Pre-College Computing Activities: A Gap Analysis. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, SIGCSE '18, pages 964–969. ACM.

Milne, L. R. and Ladner, R. E. (2018). Blocks4All: Overcoming accessibility barriers to blocks programming for children with visual impairments. *Conference on Human Factors in Computing Systems - Proceedings*, 2018-April:1–10.

Morrison, C., Villar, N., Thieme, A., Ashktorab, Z., Taysom, E., Salandin, O., Cletheroe, D., Saul, G., Blackwell, A. F., Edge, D., Grayson, M., and Zhang, H. (2018). Torino: A Tangible Programming Language Inclusive of Children with Visual Disabilities. *Human-Computer Interaction*, 00(00):1–49.

Paramasivam, V., Huang, J., Elliott, S., and Cakmak, M. (2017). Computer Science Outreach with End-User Robot-Programming Tools. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, pages 447–452.

Paspallis, N., Polycarpou, I., Andreou, P., Antoniou, J., Kaimakis, P., Raspopoulos, M., and Terzi, M. (2018). An experience report on the effectiveness of five themed workshops at inspiring high school students to learn coding. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*, pages 105–110.

Pasternak, A. (2016). Contextualized Teaching in the Lower Secondary EducationLong-term Evaluation of a CS Course from Grade 6 to 10. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, pages 657–662. ACM.

Patel, R., Breton, P., Baker, C. M., El-Glaly, Y. N., and Shinohara, K. (2020). Why software is not accessible: Technology professionals' perspectives and challenges. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–9.

Powell, N., Moore, D., Gray, J., Finlay, J., and Reaney, J. (2004). Dyslexia and learning computer programming.

Price, T. W. and Barnes, T. (2015). Comparing Textual and Block Interfaces in a Novice Programming Environment. In *Proceedings of the 12th Annual International Conference on International Computing Education Research*, pages 91–99. ACM.

Rad, M. S., Martingano, A. J., and Ginges, J. (2018). Toward a psychology of Homo sapiens: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences of the United States of America*, 115(45):11401–11405.

Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., Millner, A., Rosenbaum, E., Silver, J., Silverman, B., and et al. (2009). Scratch: Programming for all. *Commun. ACM*, 52(11):60–67.

Schanzer, E., Krishnamurthi, S., and Fisler, K. (2018). Creativity, customization, and ownership: Game design in bootstrap: Algebra. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, pages 161–166.

Schlesinger, A., Edwards, W. K., and Grinter, R. E. (2017). Intersectional HCI: Engaging Identity Through Gender, Race, and Class. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing*

589    *Systems*, pages 5412–5427. ACM.

590    Searle, K. A. and Kafai, Y. B. (2015). Boys' Needlework: Understanding Gendered and Indigenous
591    Perspectives on Computing and Crafting with Electronic Textiles. In *Proceedings of the Eleventh*
592    *Annual International Conference on International Computing Education Research*, ICER '15, pages
593    31–39. ACM.

594    Smith, N., Sutcliffe, C., and Sandvik, L. (2014). Code Club: Bringing Programming to UK Primary
595    Schools Through Scratch. In *Proceedings of the 45th ACM Technical Symposium on Computer Science*
596    *Education*, SIGCSE '14, pages 517–522. ACM.

597    Srikant, S. and Aggarwal, V. (2017). Introducing data science to school kids. In *Proceedings of the 2017*
598    *ACM SIGCSE Technical Symposium on Computer Science Education*, pages 561–566.

599    Stroub, K. J. and Richards, M. P. (2013). From Resegregation to Reintegration: Trends in the Racial/Ethnic
600    Segregation of Metropolitan Public Schools, 1993–2009. *American Educational Research Journal*,
601    50(3):497–531.

602    Swidan, A. and Hermans, F. (2017). Programming Education to Preschoolers: Reflections and Ob-
603    servations from a Field Study. In *Proceedings of the 28th Annual Workshop of the Psychology of*
604    *Programming Interest Group*.

605    Swidan, A. and Hermans, F. (2019). The Effect of Reading Code Aloud on Comprehension: An Empirical
606    Study with School Students. In *CompEd*.

607    Thomas, J. O. (2018). The Computational Algorithmic Thinking (CAT) Capability Flow: An Approach
608    to Articulating CAT Capabilities over Time in African-American Middle-school Girls. In *Proceedings*
609    *of the 49th ACM Technical Symposium on Computer Science Education*, SIGCSE '18, pages 149–154.
610    ACM.

611    Tsan, J., Rodriguez, F. J., Boyer, K. E., and Lynch, C. (2018). "I Think We Should...": Analyzing
612    Elementary Students' Collaborative Processes for Giving and Taking Suggestions. In *Proceedings of*
613    *the 49th ACM Technical Symposium on Computer Science Education*, SIGCSE '18, pages 622–627.
614    ACM.

615    Van Mieghem, A., Verschueren, K., Petry, K., and Struyf, E. (2020). An analysis of research on
616    inclusive education: a systematic search and meta review. *International Journal of Inclusive Education*,
617    24(6):675–689.

618    Van Wart, S. J., Vakil, S., and Parikh, T. S. (2014). Apps for Social Justice: Motivating Computer Science
619    Learning with Design and Real-world Problem Solving. In *Proceedings of the 2014 Conference on*
620    *Innovation & Technology in Computer Science Education*, ITiCSE '14, pages 123–128. ACM.

621    Wang, J. and Hejazi Moghadam, S. (2017). Diversity barriers in k-12 computer science education:
622    structural and social. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer*
623    *Science Education*, pages 615–620.

624    Wing, J. M. (2006). Computational Thinking. *Commun. ACM*, 49(3):33–35.

625    Yang, S., Domeniconi, C., Revelle, M., Sweeney, M., Gelman, B. U., Beckley, C., and Johri, A. (2015).
626    Uncovering Trajectories of Informal Learning in Large Online Communities of Creators. In *Proceedings*
627    *of the 2nd ACM Conference on Learning @ Scale*, pages 131–140.

628    Zhi, R., Lytle, N., and Price, T. W. (2018). Exploring Instructional Support Design in an Educational
629    Game for K-12 Computing Education. In *Proceedings of the 49th ACM Technical Symposium on*
630    *Computer Science Education*, pages 747–752. ACM.

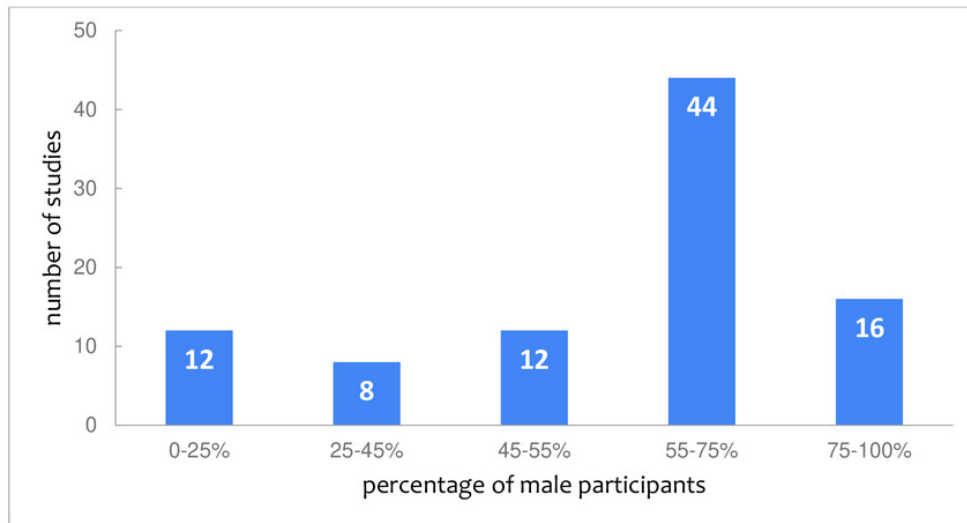
Number of papers analyzes in this paper per conference.

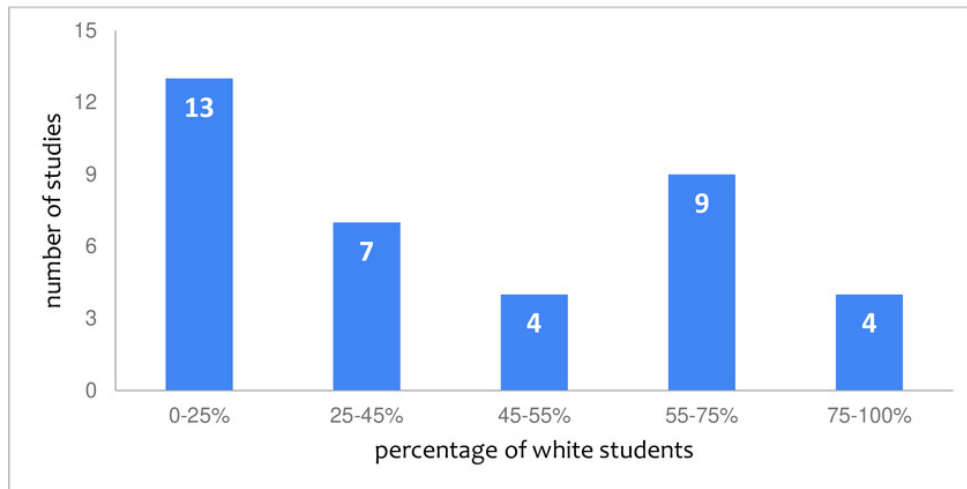
| Conference | Phase 1 | Phase 2 |
|------------|---------|---------|
| ICER | 125 | 17 |
| ITiCSE | 267 | 25 |
| SIGCSE | 561 | 90 |

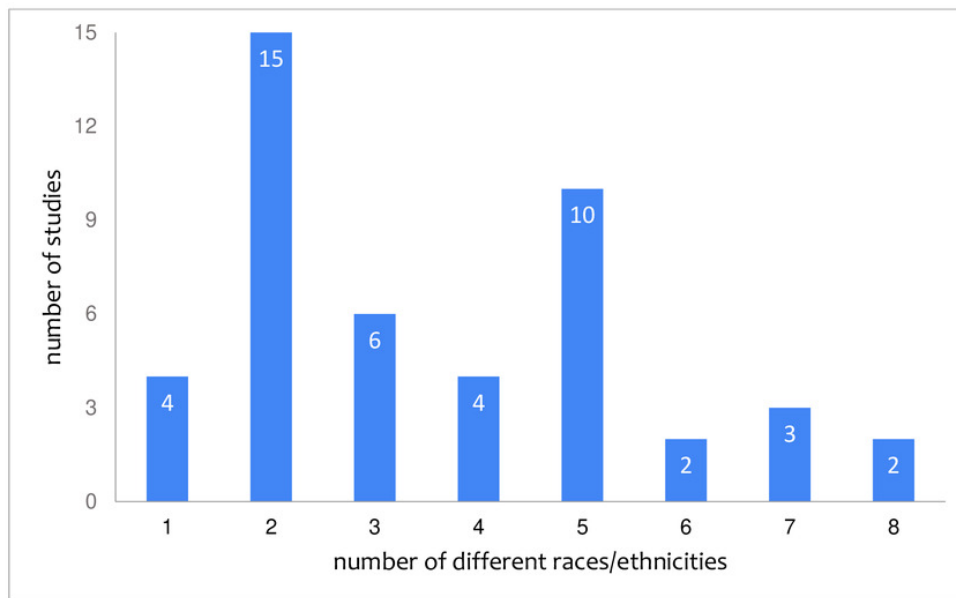
Number of studies (out of 143) that report on the different demographic categories of K-12 participants.

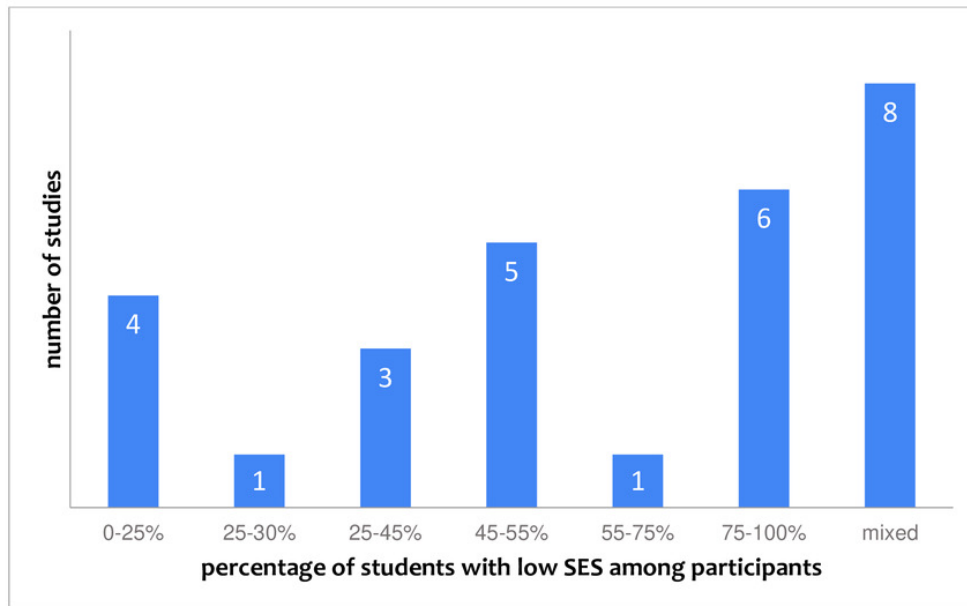
| Demographic category | Number of Studies | Percentage |
|---|---|---|
| Age/grade | 134 | 94% |
| Gender | 92 | 64% |
| Race/ethnic background | 46 | 32% |
| Socio-economic status | 28 | 20% |
| Disabilities | 6 | 4% |
| Location | 96 | 67% |
| Prior experience | 52 | 37% |

# Figure 1

Studies categorized by the reported percentage of male students.

# Figure 2

Studies categorized by the reported percentage of white students.

# Figure 3

Studies categorized by the reported number of different races/ethnicities.

# Figure 4

Studies categorized by the reported percentage of students with low socio-economic status (28 studies in total).

# Figure 5

Studies categorized by the reported percentage of students with experience in programming (52 studies in total).