# Invisible Women in IT: Examining Gender Representation in K-12 ICT Teaching Materials

Ingrid Maria Christensen*
IT University of Copenhagen
Center for Computing Education Research (CCER)
Copenhagen, Denmark

Melissa Høegh Marcher*
IT University of Copenhagen
Center for Computing Education Research (CCER)
Copenhagen, Denmark

Nanna Inie
IT University of Copenhagen
Center for Computing Education Research (CCER)
Copenhagen, Denmark

Claus Brabrand
IT University of Copenhagen
Center for Computing Education Research (CCER)
Copenhagen, Denmark

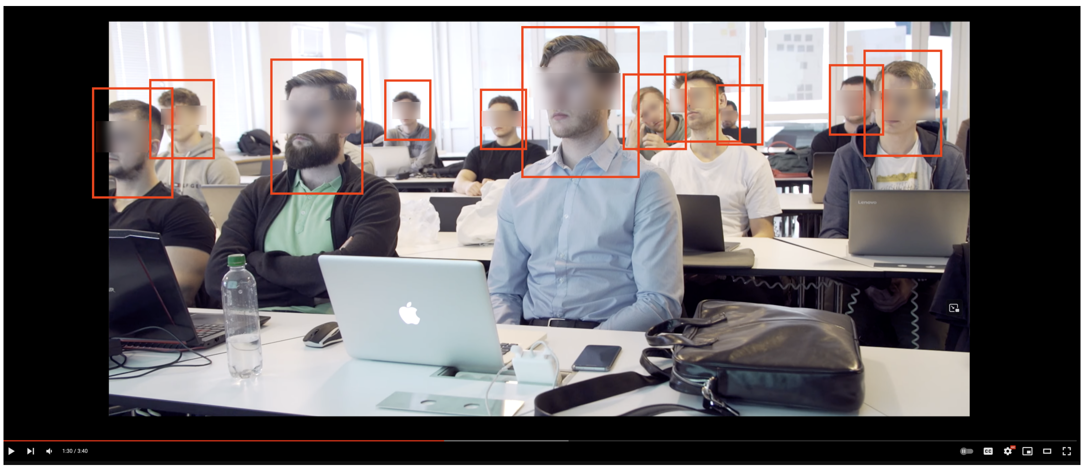


**Figure 1: Counting gender presence in audio/video material (Youtube video from KEA, Københavns Erhvervsakademi) [44].**

## ABSTRACT

The underrepresentation of women in IT has been a long-standing issue. Research has shown that women's first experience with computing education is vital in shaping their perception of the field and their perception of who "belongs" to it. In this Danish case study, we examine gender representation in the subject *Informatics*, which is the first mandatory, formal introduction to computing education for many students in Denmark. Across 25% of all Danish HHX high schools, we assess the teaching materials on parameters of gender representation in *pronouns*, *names*, *images*, *audio/video material*, and *potential role models*. Our results suggest an overall underrepresentation of women in textual and audio/video material, but not in images. We also find an overall absence of female role models, and a tendency for the materials to feature *younger* female role models than their male counterparts. We conclude with concrete recommendations for authors and publishers of teaching materials for ICT on the high school level.

*Both authors contributed equally to this research.


*ICER '24 Vol. 1, August 13–15, 2024, Melbourne, VIC, Australia*

ACM ISBN 979-8-4007-0475-8/24/08.
https://doi.org/10.1145/3632620.3671111

## CCS CONCEPTS

• **Social and professional topics** → **Computing education**.

## KEYWORDS

Teaching/learning activities, student-learning perspective, teaching design, lecture, devolution, exercise class, and institutionalization.

**ACM Reference Format:**
Ingrid Maria Christensen, Melissa Høegh Marcher, Nanna Inie, and Claus Brabrand. 2024. Invisible Women in IT: Examining Gender Representation in K-12 ICT Teaching Materials. In *ACM Conference on International Computing Education Research V.1 (ICER '24 Vol. 1), August 13–15, 2024, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3632620.3671111

## 1 INTRODUCTION

K-12 teachers have an enormous influence on students' experience of a subject. For every textbook chapter, every assignment, and every educational video in the curriculum, the teacher is not only choosing academic content. Consciously or unconsciously, she/he is also choosing the students' first formal introduction to the subject field. These choices affect students. In this exposition, the materials chosen by the teacher create an image of the subject field — in this case, the world of computing. More specifically, the materials contribute to the students' perception of who is part of the world of computing, who plays central roles, and ultimately, their perception of whether or not they fit into this world. At the end of the school year, some students will easily be able to see themselves entering the IT industry, and other students will find it harder. There are many reasons why a student can feel included or excluded, but the first introduction to the topic is heavily formative for an individual's sense of belonging.

This case study explores gender representation in teaching materials from Danish HHX high schools in the year 2022/2023. It builds on the thought that if we want to change the persistent gender imbalance in the IT sector, we need to look at the first introduction students get to the world of IT.

### 1.1 An industry lacking women

The underrepresentation of women in IT is a persistent issue, particularly in developed countries. As of 2022, women held 19% of ICT (Information and Communications Technology) jobs in Europe. This shows a mere increase of 2% from 2012 [26]. This slow rate of improvement highlights the need for even more focus on improving the gender imbalance in IT.

The absence of women in IT has implications for both individual companies and societies at large. The European Commission consistently highlights a persistent labor shortage in ICT specialist occupations [27]. This shortage is also evident in European companies' struggles with recruitment. In 2023, a staggering 63% of hiring enterprises reported difficulties in filling these vacancies [28]. Denmark faces the same challenges, with a projected demand of 22,000 ICT specialists by 2030 [41]. Failure to effectively close or significantly reduce the gender gap within IT will sustain this severe shortage of ICT specialists. On the other hand, increasing the ICT talent pool could potentially help the growing demand for IT specialists.

Meeting the need for more IT specialists is not the only reason to include women in the IT workforce. Several studies have demonstrated how diversity within teams leads to the development of better products [37, 59]. Of course, diversity encompasses not only gender, but all forms of diversity such as race, age, and educational background. Diverse teams produce better output with reduced overall bias, which ultimately translates into economic advantages for the company. A Danish report from 2016 shows that companies with higher diversity earn 5.7 percentage points more than their competitors [7], and a similar result in a report by Intel and Dalberg states how achieving representational gender diversity in leadership could help drive $320-$390 billion in increased market value [40]. Equivalently, a Boston Consulting Group report shows a positive and highly significant relationship between gender diversity and innovation in companies [78].

Additionally, the field of IT plays an increasingly crucial role in modern society. Despite financial instability in Europe in the past few years due to war and high inflation, the IT sector is one of the highest-paid industries in Denmark with record high employment rates [42, 43, 48]. IT fields offer well-paying jobs with a high level of job security, which should be equally accessible to women.

Early exposure to computing is among the many interventions shown to increase interest and aspirations for women in computing [2, 81, 82]. A 2017 report from Microsoft Philanthropies shows that most young European women become interested in STEM subjects between the ages of 11 and 12, but that interest drops significantly between 15 and 16, with limited recovery [61]. Danish and Swedish reports show a similar decrease in girls' interest in STEM between primary and lower secondary education [77]. Encouragement from parents and teachers, hands-on experience with technology, and defying computer scientist stereotypes need to happen early; before young women are discouraged from pursuing a computing interest further [77]. Ideally, initiatives to increase diversity would be implemented in existing computing education in primary schools (around the ages of 11 and 12, as the reports state).

In this paper, we examine teaching materials used in first-year high school Informatics at HHX (the Higher Commercial Examination Program) for two reasons: First, HHX offers Informatics as a *mandatory* course, not as an elective, and second, the gender distribution at HHX is more equal than at more technical high school programs in Denmark — we present the case in more detail in section 3.2. It should be noted that the official learning objectives of Informatics look similar across all different curricula in high school programs in Denmark [16], which allows us to generalize our results to some extent from HHX to the other programs. We do not intend to assess the learning outcomes of Informatics or to evaluate the efficacy of its teaching methodologies. Rather, this study's primary objective is to examine the extent to which the teaching materials currently used in Informatics align with the established recommendations advocating gender diversity and inclusion as documented in the literature.

Denmark as a case study presents a straightforward example of a Western, highly developed country where gender equality is generally high, but the ICT gender equality paradox is stark [38]. Danish is a low resource language, spoken by less than 6M people, and the results of this study are likely to generalize to similar contexts (particularly highly developed, European countries) because many teaching materials are "borrowed" from English sources.

## 2 BACKGROUND AND RELATED WORK

In this work, we review teaching curricula from two perspectives: *gender representation* and the presence of (gendered) *role models*. In the following section, we will present some of the most relevant literature and related work to these perspectives.

It is essential to note that existing literature in the topic of gender representation and inclusivity primarily discusses gender within a binary framework, lacking exploration beyond students identifying as women/men. Regrettably, this approach does not encompass the full spectrum of genders, and we must acknowledge potential limitations in directly applying our findings to other genders.

### 2.1 Gender representation

> *"The visible presence of women in the field will provide encouragement for women considering majoring*

> *in computer science as well as broaden student perceptions about the options they can pursue with a degree in computer science."* [66]

Numerous studies have consistently revealed a prevalent gender bias within high school textbooks, demonstrating a marked disparity in the portrayal of women and men [50, 52, 63]. In both textual and visual contexts, men tend to have greater representation. Moreover, depictions of women are often associated with a limited and stereotyped set of activities and careers [5, 13, 63]. Non-stereotypical gender representation is important. Especially when presented in teaching materials provided during formal education, as what young people are presented with in school is part of setting the tone for society. In computing education, among many other fields, it is especially important that students feel their gender accurately represented in order to create a sense of belonging to the field, given the gender imbalance in the IT industry generally.

Gender representation has been explored almost exclusively in textbooks. As gender presentation is a complex thing to measure, methods also vary between studies. Most often, studies utilize a mixed approach with both qualitative and quantitative methods. Quantitative approaches include finding the ratio of female to male characters (written and visually represented) [9, 53, 54], counting gendered words (pronouns, indicators of gender e.g. "girl", and gender-marked words e.g. "police-man") [51, 54, 55], generic constructions (whether a text uses generic "he") [9, 53], and order of mention (whether it is more often male-first or female-first) [53, 54]. Qualitative approaches include adjective analysis (how people of different genders are described) [9, 51], conversation analysis (who speaks first and who speaks more) [55, 84], analysis of social and domestic gender roles and heteronormativity [9, 53, 54, 68, 84], and analysis of jobs and activities [9, 53–55, 84]. Many of these studies examined English language textbooks, where the use of fictional situations with fictional characters is frequent, hence it makes sense to analyze descriptions of characters and social roles.

In some cases, genderless pronouns and titles are used instead of gendered ones. This can be done in English (and Danish) by using the singular *they* as a generic pronoun and by using titles such as"teacher" and "programmer". Although this approach is inclusive of gender identities outside the binary woman/man, genderless language is not necessarily the most gender equal. Studies have shown that using gender-symmetrical terms, like the generic *he/she*, led to greater visualization of female actors compared to genderless terms, like the generic singular *they* [39, 76]. Even genderless terminology (like *chairperson*) may lead to a gendered interpretation more often than gender-symmetrical terms (like *chairman/woman*) [65]. In a field like computing, that is so strongly tied to masculinity and male stereotypes [82], genderless terms with possible gendered interpretations are critical to consider. For example, a "programmer" with a non-specified gender may in more cases be interpreted as male. As Perez states when speaking of genderless language in her book, *Invisible Women*; "*because men go without saying, it matters when women literally can't get said at all*" [64].

## 2.2 Role models

> *"Women tend to find role models more important than men, and especially to value having access to role models when they participate in a male-dominated area such as engineering and computing."* [74]

Role models can inspire and influence individuals to aspire toward certain traits or accomplishments, guiding them in making positive choices and shaping their attitudes, beliefs, and actions. To young women in computer science, role models may be successful women in tech, online educators and influencers, older students, peers, parents, or teachers. Several studies argue that access to role models in computer science is vital in recruiting the next generation of women [61, 66, 74, 77, 83]. A Plan International survey found it to be the most influential factor in encouraging girls and young women to pursue the STEM fields [77].

Role models play a pivotal role in reshaping and breaking stereotypes. Several studies advocate for a broad range of role models to show the field's diversity and boost women's interest [14, 33]. In particular, they argue that underrepresented students will likely feel more engaged in computing if they see role models in computing who value aspects beyond the technical. Exposure to female role models who share similarities with oneself can significantly alter implicit self-views, enabling young women to envision themselves in the computing field [8]. However, it is crucial to note the nuanced impact of role models; reinforcing geek stereotypes through female role models might not effectively enhance computer self-efficacy among female non-majors in computer science [20].

Teachers are essential role models in fostering interest and confidence among female students in the field of computer science. One study found teachers to be key inspirations in students' educational choices [71], while another specifically tied female computer teachers to a higher inclination in female students to pursue computer science majors at university [12]. At Stanford University, offering many positive role models for women has been one of the factors in increased enrollment in the introductory computer science course [66]. While teachers as role models are important for recruitment, they are equally important for retention in the field. In a study of engineering majors, exposure to effective female role models within a university department was significantly related to the intent to remain in the major and affected long-term goals, particularly for women [4].

The impact of role models with varying genders, backgrounds, races, and so on in teaching has been explored in numerous studies [17, 46, 72]. Also specifically within IT, confirming the effect on the choice of education, among other things [3, 33]. Other studies have explicitly examined the specific effect of role models of the same gender as the students [10, 58], concluding that female role models affect female students to a higher degree than male role models. Both studies comment on whether the same effect is true for male students, one concluding the same effect is present but only in female-dominated sectors [10], while the other says it requires further exploration [58].

## 2.3 Omitted relevant perspectives

While we could have included numerous different aspects in our analysis, certain perspectives have been omitted from this study. Although important and relevant for diversity and inclusion in computer science, they lay outside the scope of this specific project.

We will briefly present four omitted perspectives and explain our reason for excluding them.

*2.3.1 Collaboration.* Collaboration in computing education has repeatedly proven to be an effective form of learning. Not only has it demonstrated a significant positive effect on knowledge gain, skill acquisition, and student perception [19], but it also mirrors real-world scenarios, aiding in preparing students for future careers [69]. Additionally, studies have shown the profound impact of collaborative learning on both class retention and students' inclination to pursue further STEM courses. Notably, a study shows how students, especially female students, are more likely to drop out of a physics course taught in a traditional lecture style rather than a mixed style including collaboration [29]. Moreover, collaborative approaches in computer science can help in dispelling stereotypes of anti-social computer scientists working in solitude and can showcase the field as helpful, altruistic, and community-oriented [14]. The collaboration perspective is not part of our study for two reasons. The main reason is that the scope of the study is *teaching materials* and not classroom teaching. We therefore do not consider how students have worked with the materials in practice. Additionally, social relations between students are not established in Informatics classes alone, as Danish high school students attend nearly all the same classes for the duration of their high school education.

*2.3.2 People vs. Things.* In 1998, Richard A. Lippa discovered a correlation between gender and vocational interests in the People-Things dimension [56]. Lippa conducted three studies, consistently revealing a significant relationship between women preferring People, and men preferring Things. The findings have later been confirmed both by Lippa self in a 2010 study, concluding that "*Men tend to be much more Things-oriented and much less People-oriented than women*" [57], but also in other studies [67, 75]. Two distinct studies, conducted in Denmark specifically, investigated the correlation between gender and the People-Things dimension within computing education. In the first study [21], involving 500 high school students, woman chose People-themed tasks 2.7 times more often than men. Furthermore, for those without prior programming experience, the odds of choosing People-centric tasks were 1.4 times higher than that of people with previous experience. Combining these effects, the odds of women without prior programming experience favoring People was 3.8 times higher than for men with programming experience. The second study [60] examined 152 first-year university students' preference for either People- or Things-themed mandatory assignments. The results showed that first-year female university students tended to favor People-themed computing tasks over Things-themed tasks, although the two tasks had the exact same educational content. Another study, analyzing university programs' advertisement material and courses, showed how favoring People in these materials appealed more to female students [32]. As in the case of collaboration, this perspective is particularly relevant to practical work in the classroom, which has not been the focus of this study.

*2.3.3 Relatability.* Relatability of computing teaching material has proven to be one of the factors attracting underrepresented students to computer science [31]. A Microsoft report states that "girls become more interested in STEM once they're able to conceive what they can do with these subjects, how they can be applied to real-life situations, and how they might be relevant to their future" [61]. It is easier for students to see the relevance of computing if they are presented with materials that resonate with their lives [35], and unfortunately, many computer science courses fail to make the connection between computing concepts and their application to technologies students use day-to-day [31]. Relatability as a perspective has not been analyzed, as it has not been shown to improve gender diversity specifically, but create better computing education more generally.

*2.3.4 Programming language and syntax.* Programming language and syntax have shown to have an effect on the appeal of programming and the confidence of novice programmers. Some programming languages, such as Java and C++, have been criticized for being too complex for beginners. It has been argued that these languages' syntax, semantics, etc. can distract beginners from learning basic programming concepts, which may have an impact on students' confidence in their programming abilities [47]. As this perspective is programming-specific, it could only have been applied to the fraction of our data concerning programming education. In addition, it is not gender specific, meaning that programming language and syntax have an effect on novice programmers versus experienced programmers, but not on students of different genders.

*2.3.5 Teaching language.* The language in which computing education is taught is an important factor in a student's learning. Except for a few exceptions like Scratch [70] and Hedy [36], which have multiple language options, almost all programming languages are based on the English language. This is a barrier for non-native English speakers. One study showed that students enrolled in an English version of a course otherwise in Swedish (the native language in the study) answered statistically significantly fewer test questions correctly [15]. Another study specific to computer science found that non-native English speakers have to work harder to achieve the same grade as their native English-speaking peers [34]. Although the teaching language in Danish high schools is always Danish, teachers may include English language teaching materials. Despite being a factor in the students' learning, teaching language is not a factor in gender inclusion. For this reason, the perspective was not considered in the analysis.

## 2.4 Gender-inclusion in a Danish educational context.

Focus on gender equality and combating gender stereotypes remains limited in the Danish public school system [25, 49]. In 2022, the Danish Union of Teachers commenced a project on gender awareness for teachers in primary and lower secondary education [24]. GL, an association for teachers in high school, has also expressed their wish to increase gender equality and combat gender stereotypes in all high school programs [30]. Looking specifically at gender balance within teaching materials, one project found a stark gender imbalance in the 10 most used history books in Danish schools [6]. Unfortunately, this is the only example we were able to find on existing teaching materials in Denmark. To the best of our knowledge, our study is the first to assess gender representation in Informatics teaching material in Denmark or other countries.

# 3 METHODOLOGY

## 3.1 Objective

Given the proven impact of early exposure on an individual's likelihood of considering a career within IT, the aim of this study was to explore the earliest mandatory introduction to computing in the Danish school system. The study focused on the following research questions:

> **RQ1 (Gender representation):** To what extent do Danish K-12 Informatics teaching materials portray equal *gender representation*?

> **RQ2 (Role models):** To what extent do Danish K-12 Informatics teaching materials portray equal gender representation of *role models*?

## 3.2 Case description: HHX schools in Denmark

As of 2024, computing (or Informatics) is still not taught as a mandatory subject in Danish primary schools (despite recent efforts of introducing a new school subject to be entitled *"technology comprehension."*) After nine years of compulsory primary school, many children continue to high school. In Denmark, there are four types of high schools: STX (Higher General Examination Program), HF (Higher Preparatory Examination), HTX (Higher Technical Examination Program), and HHX (Higher Commercial Examination Program). They all last three years (although HF is only two years, but requires one more year of primary school). STX is the most popular choice with 56% of high school students enrolled, but Informatics is only offered as an *elective* course on that program. We have thus chosen to focus our study on HHX (Higher Commercial Examination Program) which focuses on subjects related to the fields of business economics, marketing, and international economics. We chose this K-12 program because Informatics is a *mandatory* course and it is relatively gender balanced (41% girls, 59% boys). For many students, the Informatics course on the HHX program will thus constitute the first encounter with and exposure to computing education.

## 3.3 Corpus building

To build a corpus, we systematically searched through the websites of all 64 HHX schools in Denmark [79], downloading available Informatics syllabi for the 2022/2023 school year. The Danish Ministry of Children and Education specifies which teaching *topics*, but not which teaching *materials* should be included in first-year Informatics [16]. Because of this, the downloaded syllabi comprise the teaching materials individual Informatics teachers have deemed fitting for that school year. Unfortunately, several schools remove these syllabi from their websites outside of examination season, which narrows the number of publicly available syllabi. To ensure comprehensive geographical coverage, we extended our reach by contacting schools located in areas we had not yet covered. The additional syllabi were sent to us via email. This approach resulted in a total of 28 distinct syllabi from the 2022/2023 school year used in 55 different classes at 16 different schools, corresponding to 25% of all HHX schools in Denmark and national geographical coverage. This corpus constitutes the *total corpus*. When analyzing *individual syllabi* (i.e., comparing syllabi to each other), two syllabi were excluded from the total of 28, as they had only two and six teaching materials listed respectively, and would have been extreme outliers in the dataset. The *individual syllabi* corpus therefore consists of 26 unique syllabi, containing between 20–108 teaching materials each (49 on average). The appendix includes a map with the geographical distribution of schools from which we collected syllabi.

There are obviously vast amounts of possible teaching materials from books and the internet that a teacher could use. However, 41% (255/616) of the listed teaching materials were used in more than one syllabus, and the two materials with the highest usage were included in 12 different syllabi. Therefore, although an approximation, the materials do, to a large extent, reflect what is available to Informatics teachers.

#### 3.3.1 *Included materials.*

We carefully reviewed each syllabus, ensuring that all mentioned teaching materials were in our database. We classified these materials into five primary material formats: `textbook`, `audio/video`, `web page`, `assignment`, and `document`. Each 'material' constitutes the smallest unit of analysis of the study.

- `Textbook` encompasses materials from published, Danish textbooks. All textbooks mentioned in the syllabi are available as online textbooks. Textbooks are further divided into sub-chapters and sub-sub-chapters in our database to align with the structure and specificity represented in the syllabi. Each subdivision then corresponds one 'one material'.
- `Audio/video` covers a diverse range of YouTube videos, embedded videos on listed web pages, or within online textbooks.
- `Web page` includes articles and online tutorials.
- `Assignment` encompasses both teacher-created assignments and assignments available on various websites.
- `Document` includes documents explaining Informatics theory from various authors.

The systematic review and categorization of all syllabi yielded a total of 616 distinct teaching materials (see Table 1).

#### 3.3.2 *Excluded materials.*

Simultaneously, we excluded several materials for various reasons. A substantial subset of the listed materials had broken links, while others lacked sufficient specification, such as textbooks listed without references to specific chapters. Materials with complex formats, such as online `games` or `tools`, proved hard to analyze and were excluded. Additionally, we opted to filter out lengthy `audio/video` materials (>35 minutes), like movies or TV series. Some materials required payment or a specific account for access, and certain materials were intended solely for the teachers. Lastly, one teacher uploaded their `slideshows`, but since this was an exception to everyone else, we excluded these materials. In total, 123 teaching materials were excluded. Table 1 and 2 provide detailed overviews of our included and excluded materials, respectively.

#### 3.3.3 *Overview of collected data.*

Figure 2a illustrates the percentage distribution of unique materials collected. An important factor of the distribution of the gathered material, is how much each syllabus referred to each material, or, how much 'weight' each material carried in each syllabus. Figure 2b shows the weighted distribution highlights the percentage of references to each material format. The

| Format | n |
|---|---|
| textbook | 233 |
| audio/video | 222 |
| web page | 132 |
| assignment | 21 |
| document | 8 |
| **Total** | **616** |

**Table 1: Included materials.**

| Reason | n |
|---|---|
| Broken link | 52 |
| Unspecified | 29 |
| Complex format | 14 |
| Long format | 12 |
| No access | 7 |
| For the teacher | 5 |
| Slideshow | 4 |
| **Total** | **123** |

**Table 2: Excluded materials.**

graph emphasizes the significant use of textbooks compared to the other formats. When examining these figures, note that material sizes vary.

## 3.4 Data analysis

For all materials, we logged fundamental information used for classification purposes and subsequent analyses. The properties used to collect the information are presented in Table 3. The `name` property is the name of the material. It works as a unique identifier used to avoid duplicates in the dataset. The `format` is an Enum corresponding to the five general format categories mentioned earlier. Each material is also mapped to the specific syllabus(es) where it came from, indicated in the `syllabus` property. The `language` property indicates the content language. Lastly, the `source` property indicates the author or publisher of the material, for example, "Code.org".

*3.4.1 Gender representation.* For research question 1, we evaluate gender representation on three parameters: usage of female and male *pronouns*, occurrences of female and male *names*, and *amount* of women and men in images, audio, and video. Note, that while it is grammatically correct to write *female* and *male* pronouns and names, we will hereafter use the terms *women's* and *men's* pronouns and names, to keep the terminology consistent when counting *women* and *men* in images, audio, and video.

While the initial intention was to include genderless pronouns in the count, we ultimately decided not to do so. In Danish, the non-gendered pronouns "de" and "dem" ("they" and "them") are also used as articles indicating the definite plural form of a noun (for instance, "*[de] mest besøgte hjemmesider.*" — ("*[the] most visited websites.*") To accurately count instances of "de" and "dem" only when used as pronouns, a contextual analysis for each occurrence would be necessary. We streamlined the process by utilizing a context-insensitive search function without having to interpret the context. It is worth noting that we came across only two instances of the use of a unisex name (*Jackie* and *Taré*), and for the same reason, we decided to exclude unisex names from our dataset. Further, we encountered multiple uses of the gender-inclusive use of "he/she" (e.g., "*[...] hvilken strategi [han] eller [hun] vil benytte [...]*" — "*[...] which strategy [he] or [she] will utilize [...]*"). From these observations, we hypothesized that the utilization of the genderless pronouns "de" and "dem" ("they" and "them") are not frequently used in Danish teaching materials.

For audio/video material, we counted the amount of individuals featured in the video and noted their gender, rather than attempting to create some arbitrary unit of measurement for the frequency of an individual's presence in the video. Evaluating by frame would be excessively detailed and hinder comparison across formats or videos of varying lengths. Alternatively, assessing by cut would also vary considerably from one video to another. Consequently, we decided to exclusively count distinct individuals' genders, particularly because individuals prominently featured in the materials are considered and analyzed in the *role model analysis*. In text material, we counted both instances of pronouns, names, and gender in images, while in `audio/video` material, we only counted people's gender. Figure 3 shows this flow. We hypothesized that the visual presence of someone would carry more importance in a video than their potential mentioning of other people. A content analysis might add more insights and perspectives to this work, but it is out of the scope of the current study.

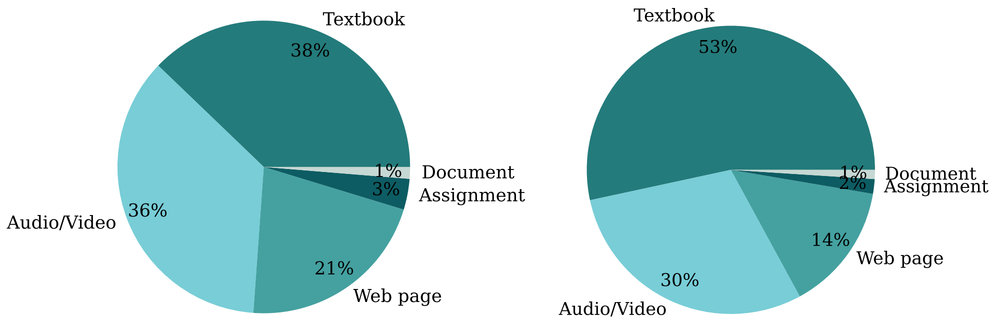
Table 4 shows the properties utilized in the gender representation analysis. All properties are counted in integers. `w` is short for women and `m` for men.

*Examples.* Figure 4 shows an example from a textbook. Pronouns are highlighted in blue, while names are highlighted in orange. The analysis for this case would yield: `w_pronouns = 1, m_pronouns = 2` and `w_names = 1, m_names = 1`. Similarly, Figure 5 displays a screenshot from a YouTube video featuring two individuals. In this example, the analysis would result in: `w_audio-video = 1, m_audio-video = 1`.

*3.4.2 Data categorization.* We use the categories "Mostly women", "Equal", and "Mostly men" to describe teaching materials when answering RQ1. The categorization relies on strict comparison. For instance, having two women and one man in a teaching material would still be classified as "Mostly women". As shown in Table 5, the data in this study deals with relatively small margins (an average less than four mentions of a gendered entity), and we therefore believe the strict comparison categories are sensible.

*Statistical analysis.* We present the statistical significance of the gender representation analysis by finding the probability of the observed outcome from a binomial distribution. Because of the complex nature of our data, it does not live up to the requirements the data has to meet to calculate binomial probability. For that reason, the significance calculated should not be standalone results but rather viewed as supplementary to the graphs. Before calculating significance, we make three overall assumptions for all four gender representation properties. First, materials are independent of each other in terms of the property we measure (e.g. the pronouns counted in one material do not affect the number of pronouns counted in another material). Second, we exclude the materials with an equal distribution of women and men for that property to have only two categories. Third, we assume there is a 50% chance of getting a material with either "mostly women" or "mostly men" for that property. In all calculations, we use a significance level of 0.05.

*3.4.3 Role models.* Whereas the *gender representation analysis* gives an overview of all people represented in the materials, the role




**(a) Unique materials.** **(b) Syllabus references to materials.**

**Figure 2: Distribution of materials in reviewed syllabi overall. Figure (a) (left) shows the prevalence of different formats, while Figure (b) (right) shows how much weight each format carried in the syllabi, i.e., how often the different formats were referenced in the syllabi.**

| | | | | |
|---|---|---|---|---|
| name: string | format: enum | syllabus: number[] | language: string | source: string |

**Table 3: General properties of teaching materials.**

| | | | |
|---|---|---|---|
| w_pronouns: int | w_names: int | w_images: int | w_audio-video: int |
| m_pronouns: int | m_names: int | m_images: int | m_audio-video: int |

**Table 4: Properties of the *Gender representation analysis.***

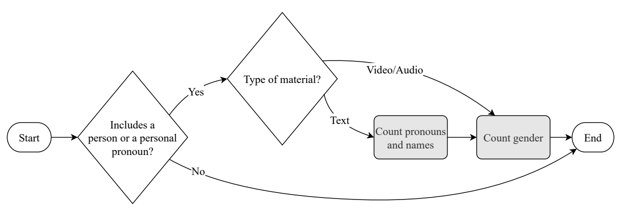



**Figure 3: Flowchart showing *gender representation analysis.***

| | Mean ± SD | Range |
|---|---|---|
| w_pronouns | 2.4 ± 5.2 | [0 .. 29] |
| m_pronouns | 2.6 ± 2.8 | [0 .. 13] |
| w_names | 2.3 ± 4.7 | [0 .. 38] |
| m_names | 3.9 ± 5.2 | [0 .. 40] |
| w_images | 1.5 ± 1.6 | [0 .. 8] |
| m_images | 1.6 ± 1.9 | [0 .. 9] |
| w_audio-video | 1.7 ± 2.7 | [0 .. 15] |
| m_audio-video | 2.6 ± 4.5 | [0 .. 39] |

**Table 5: Statistical summary for each property in the Gender Analysis dataset: mean, standard deviation, minimum, and maximum.**

model analysis considers more closely the people who fill prominent roles in teaching materials. Definitions of a role model vary both in the literature and in everyday life [33]. The most common description is that a role model is someone to imitate, but the formal definitions tend to include varying additional details to the role model concept, e.g. an element of admiration, an age difference, or that the imitation is limited to a specific role [33]. As these definitions of a role model are all highly subjective, we defined three criteria for a person to be a *potential* role model in this study. First, the person is *real*. Second, *the person's face can be seen* in a photo or video. Third, we are explicitly told that *the person has something to do with IT*. Examples of the last criterion include, for instance, computer science students, IT project managers, and big data experts. After going through all the prominent people in the materials (meaning every person presented with name, gender, *and* additional information about her/him), we filtered out any who did not meet the criteria as a potential role model. Figure 6 shows an example of a person who has been included as a potential role model and a person who has been excluded. The second person was excluded since we cannot see the person's face.

*Statistical analysis.* When calculating the probability of some gender in potential role models from a binomial distribution, we make two assumptions. First, we assume that the gender of potential role models are independent data points (e.g. the gender of one potential role model does not affect the gender of another). Second,

| gender: string | traits: enum[] | role_model: bool | description: enum[] |
|---|---|---|---|

**Table 6: Properties of the *Stereotype and Role Model Analysis*.**

**Meaningful play**

Meaningful play is what occurs when the relationships between actions and outcomes in a game are both discernable and integrated into the larger context of the game.

Katie Salen og Eric Zimmerman

Hvis spil for alvor skal være interaktive, må spilleren også selv have lov at vælge hvilken strategi han vil benytte til at nå frem til sit mål. Og lader man spilleren vælge selv, kan den ene strategi ikke være bedre end den anden.

Det kan vi ved hele tiden at kommunikere klart til spilleren, om han eller hun er på rette vej. Det kan ikke nytte, at man bliver ved med at slå på en tilsyneladende helt upåvirket modstander, hvis modstanderen netop skal bekæmpes med slag.

**Figure 4: Text material (constructed example from a textbook). Pronouns are highlighted in blue and names are highlighted in orange. The bottom text translates: *"If games are truly interactive, the player must be able to choose the strategy he wants to utilize to reach a goal. And if one lets the player choose their own, one strategy can not be better than another. We can do this by continuously communicating to the player if he or she is on the right course. It does not work that one keeps punching a seemingly unaffected opponent if the opponent is, in fact, supposed to be fought with punches." Source: Systime [62].***

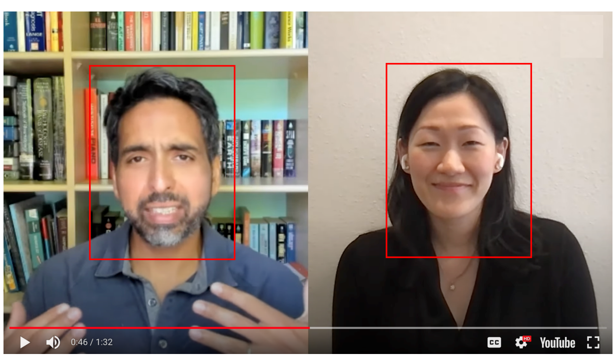



**Figure 5: Audio/video material (from a YouTube video). *Source: Khan Academy [45].***

we assume that there is a 50% chance of the role model being either a woman or a man. As before, we use a significance level of 0.05.

*3.4.4 Inter-rater reliability.* The two first authors conducted the analysis in parallel. Inter-rater reliability (IRR) was calculated initially before beginning the analysis of our teaching materials. This process consisted of four iterations, where they independently analyzed 10 teaching materials in each iteration. Each set of 10 materials included examples from all five formats, selected otherwise arbitrarily. The percent-agreement method [73] or percentage of absolute agreement method [18] was employed, involving the counting of identical analyses, which were then divided by the total number of analyses. For all analyses, they were only counted as identical if all properties were analyzed in the same way. Following each iteration, discrepancies were discussed and guideline further refined. The outcomes of the four iterations are listed in Table 7.

| Iteration | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Agreement | 54% | 90% | 88% | 94% |

**Table 7: Results from the IRR iterations.**

*3.4.5 Data processing.* The material analysis resulted in one large dataset with columns corresponding to the listed properties in Table 3, 4, and 6. Hereafter, we conducted our data processing and data analysis in a Jupyter Notebook in Google Colaboratory using `Python 3` with the data manipulation and analysis library `pandas`, the numerical computing package `NumPy`, and the plotting and data visualization libraries `matplotlib` and `seaborn`. Further, we used ChatGTP for fixing errors and in some cases for help with creating the plots.

# 4 RESULTS

Each research question is represented by a separate section. Under each section we look at both *total corpus* and *individual syllabi*, as explained in the *Data analysis* section (3.4) in the Methodology.

## 4.1 Gender representation (RQ1)

*4.1.1 Total corpus.*

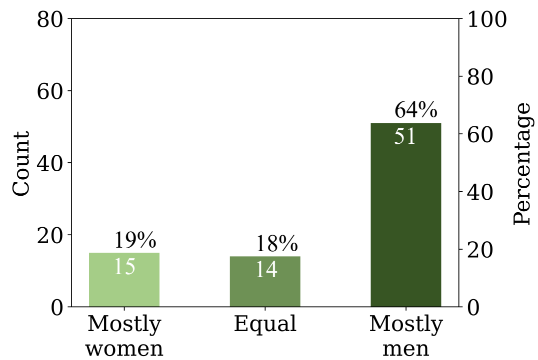
**Pronouns.** Across all materials that include pronouns (see Figure 7a), 19% of materials include mostly women's pronouns, 18% include an equal amount of women's and men's pronouns, and 64% include mostly men's pronouns — or, **nearly 2/3 of materials exhibit a majority of men's pronouns**. This represents a highly skewed distribution of pronouns in materials. The probability of an outcome $X$ of at least 51 materials with mostly men's pronouns $n$ (out of 66 materials in total) is nearly zero. It means that this distribution is nearly impossible if there is a 50% chance of "mostly women" or "mostly men". This finding is highly statistically significant.

**Names.** Across all materials that include names (see Figure 8a), 20% include mostly women's names, 8% include an equal amount of women's and men's names, and 72% include mostly men's names — or **nearly 3/4 of materials containing names exhibit a majority of men's names**. Again, we see a highly skewed distribution of names in materials. The probability of getting an outcome $X$ of at least 77 materials with mostly men's names $n$ (out of 98 materials in total) is nearly zero. It means that this distribution is nearly


**(a) Included.** ***Source: Code.org [22]***


**(b) Excluded.** ***Source: UX Mastery [80]***

**Figure 6: Example of people included and excluded as a potential role model.**

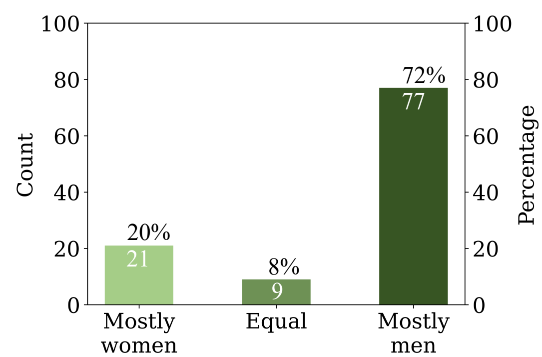



**(a) Distribution.**

| | Mostly women | Mostly men | Total |
|---|---|---|---|
| $n$ | 15 | 51 | 66 |
| % | 23% | 77% | 100% |
| $P(X \geq n)$ | 0.999 | 0.000 | |

**(b) Count, percentage, and probability.**

**Figure 7: Pronouns in total corpus.**




**(a) Distribution.**

| | Mostly women | Mostly men | Total |
|---|---|---|---|
| $n$ | 21 | 77 | 98 |
| % | 21% | 79% | 100% |
| $P(X \geq n)$ | 0.999 | 0.000 | |

**(b) Count, percentage, and probability.**

**Figure 8: Names in total corpus.**

impossible if there is a 50% chance of "mostly women" or "mostly men". This finding is highly statistically significant.

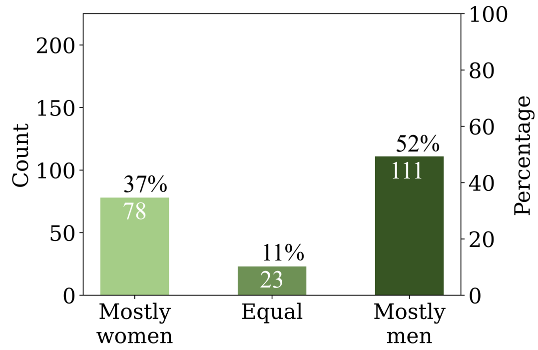
> **OBSERVATION 1:** The vast majority of our analyzed teaching materials corpus for Danish K-12 Informatics exhibit mostly men's pronouns and men's names.

**Audio/Video.** Across all audio/video materials (see Figure 9a), 37% include mostly women, 11% include an equal amount of women and men, and 52% include mostly men. This represents a slightly skewed distribution of women and men in audio/video materials. The probability of getting an outcome $X$ of at least 111 materials with mostly men $n$ (out of 189 materials in total) is around 1%. It means that this is very unlikely if there is a 50% chance of "mostly women" or "mostly men". This finding is statistically significant.

> **OBSERVATION 2:** The majority of our analyzed teaching materials corpus for Danish K-12 Informatics include mostly men in audio/video materials.

**Images.** Across all materials that include images of people (see Figure 10a), 34% include mostly women in images, 25% include an equal amount of women and men in images, and 41% include mostly men in images. This represents **a close to equal distribution of women and men in images**. The probability of getting an outcome $X$ of at least 30 materials with mostly men in images $n$ (out of 55 materials in total) is around 30%. It means that this distribution is likely if there is a 50% chance of "mostly women" or "mostly men" in images. This finding is *not* statistically significant.

> **OBSERVATION 3:** We observe a nearly even distribution of Danish K-12 Informatics materials that exhibit

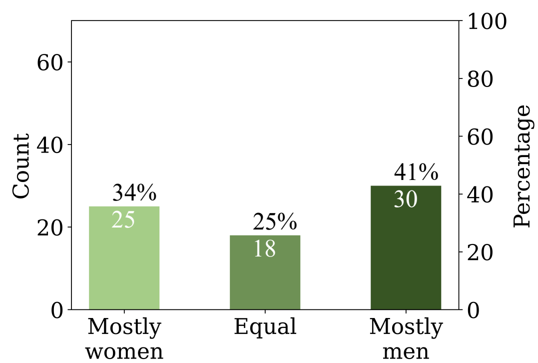



**(a) Distribution.**

| | Mostly women | Mostly men | Total |
|---|---|---|---|
| $n$ | 78 | 111 | 189 |
| % | 41% | 59% | 100% |
| $P(X \geq n)$ | 0.993 | 0.010 | |

**(b) Count, percentage, and probability.**

**Figure 9: All audio/video materials.**




**(a) Distribution.**

| | Mostly women | Mostly men | Total |
|---|---|---|---|
| $n$ | 25 | 30 | 55 |
| % | 45% | 55% | 100% |
| $P(X \geq n)$ | 0.791 | 0.295 | |

**(b) Count, percentage, and probability.**

**Figure 10: Images in total corpus.**

mostly men and materials that exhibit mostly women in images.

4.1.2 *Individual syllabi.* Figure 11 shows the distribution of syllabi on the four properties in our *gender representation analysis*: pronouns, names, images, and audio/video. Each is a distribution of the percentage of women. The figure shows that **75% of all syllabi have less than 50% women on all properties.**

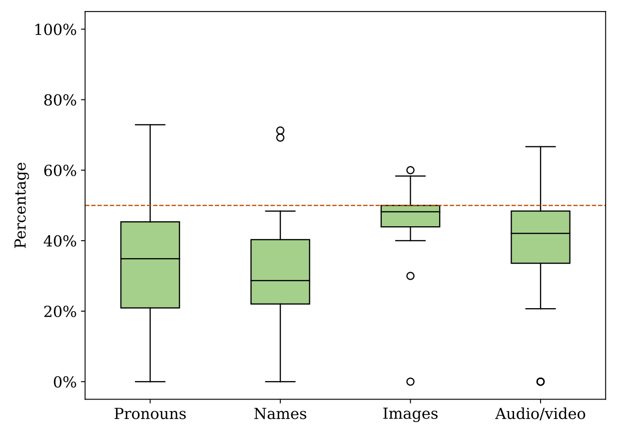
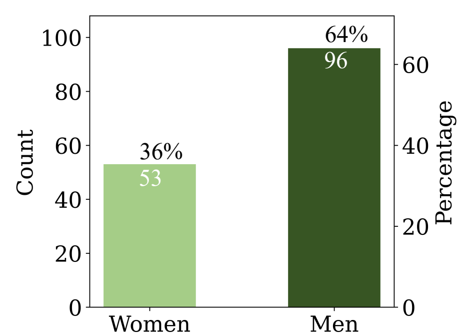



**Figure 11: The distribution of women's presence in syllabi. The orange line marks 50%.**




**(a) Distribution.**

| | Women | Men | Total |
|---|---|---|---|
| $n$ | 53 | 96 | 149 |
| % | 36% | 64% | 100% |
| $P(X \geq n)$ | 0.999 | 0.003 | |

**(b) Count, percentage, and probability.**

**Figure 12: All potential role models.**

## 4.2 Role Models (RQ2)

4.2.1 *Total corpus.* Across all mentioning potential role models (see Figure 12a), 36% are women and 64% are men. This represents a skewed distribution of gender in potential role models, as nearly 2/3 of all potential role models are men. The probability of getting an outcome $X$ of at least 96 potential role models who are men $n$ (out of 149 potential role models in total) is around 0.3%. It means that this is very unlikely if there is a 50% chance of either gender. This finding is highly statistically significant.

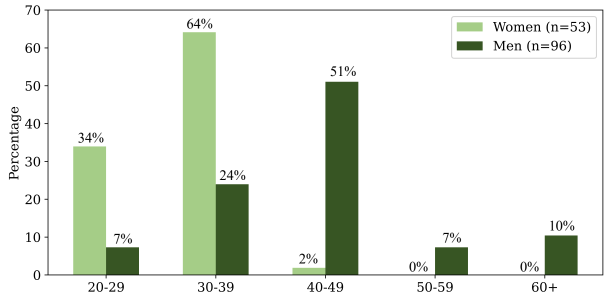



**Figure 13: Distribution of ages in all potential role models.**

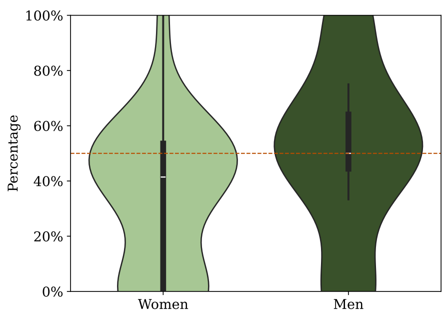



**Figure 14: The distribution of potential role models in syllabi. The orange line marks 50%.**

> <u>Observation 4</u>: Nearly 2/3 of all potential role models in our analyzed Danish K-12 Informatics teaching materials corpus are men.

We note, that we see a clear difference in ages among women and ages among men in potential role models, where **potential female role models tend to be younger than men** (see Figure 13).

*4.2.2 Individual syllabi.* Figure 14 shows a violin plot of potential role models where the black vertical rectangle and lines represent a boxplot and the width of the plot shows the frequency of syllabi with that percentage. Although the differences between the two plots are subtle, the extremes are noteworthy. At the bottom, we see a higher frequency of syllabi with 0% women than 0% men. At the top, we see almost no syllabi with 100% women, but quite a few syllabi with 100% men. Note that these plots are not symmetrical, as there are syllabi with no potential role models at all.

Table 8 shows the number of syllabi that fall into each category for each of the gender inclusion properties. We see how for each property there are syllabi with more women and even a few with significantly more women. However, most often we see syllabi with more men and to a much larger extent than for women, significantly more men. This tells us that there is a large variation between syllabi, but that men generally take up more space in syllabi than women.

> <u>Observation 5</u>: Despite variation between syllabi, women are underrepresented in a majority of the analyzed individual Danish K-12 Informatics syllabi.

# 5 DISCUSSION

## 5.1 Observation 1: The vast majority of our analyzed teaching materials corpus for Danish K-12 Informatics exhibit mostly men's pronouns and men's names.

The results show a clear underrepresentation of women's pronouns and names in text materials. This finding exposes at least two challenges. First, the idea of a default male in IT gets reinforced for the student when a substantial percentage of the materials read "he" instead of "she". Second, even when the teacher is aware of gender inclusion in their teaching, finding appropriate materials in terms of equal gender representation, proves difficult when the total corpus distribution is skewed.

The pronouns and names counted vary in context. Among others, they included people in examples, names of researchers whose models were presented, and quotes from experts in the field. Of course, the number of total pronouns and names one sees in a teaching material is not necessarily an indicator of how well that gender is represented. For example, it could be that there is an equal gender balance overall, but all the women mentioned are anonymous, fictional women in examples, while all the men are researchers and experts in the field. The analysis was done completely free of context, meaning we did not analyze the context in which individual pronouns or names were presented and therefore cannot say how gender names and pronouns were represented in or across different contexts. For future research, it would be relevant to investigate the contexts of all people represented in teaching materials, as these could either exacerbate or mitigate our results.

## 5.2 Observation 2: The majority of our analyzed teaching materials corpus for Danish K-12 Informatics include mostly men in audio/video materials.

This analysis was also conducted without considering context. In this paper, we therefore do not analyze the role a person played in the video. Much of current computing industry is male-dominated and a direct consequence of this is difficulty of finding female experts for presentation and quotes. However, although women have played integral roles in the history of computing since its inception, none of the materials from Danish textbook publisher acknowledge their contributions. When mentioning Alan Turing, a pioneer in theoretical computer science, it also makes sense to mention Ada Lovelace, who is cited by many as the first-ever computer programmer. The uneven gender distribution supports the general idea that IT is for men and not women.

## 5.3 Observation 3: We observe a nearly even distribution of Danish K-12 Informatics materials that exhibit mostly men and materials that exhibit mostly women in images.

When omitting teaching materials with an equal amount of women and men (25%), the distribution of women and men in images is 34% and 41%, respectively. It seems attainable to achieve equal

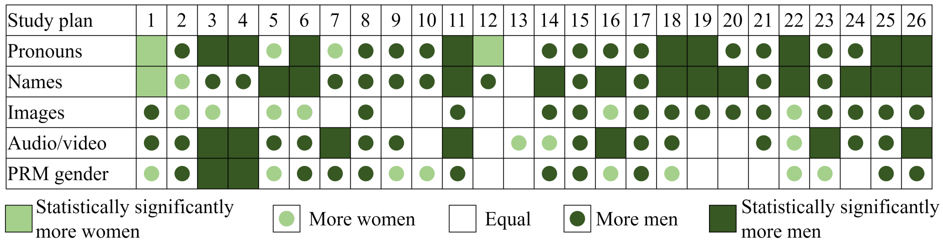



**Figure 15: Statistical significance and majorities of all gender properties in all syllabi. Each number at the top represents a syllabus, while the properties are listed vertically to the left (PRM is short for "potential role model"). For each property in each syllabus, we calculate its significance by comparing the count for women and men with the total. Significance was calculated with a probability from a binomial distribution with the assumption that all properties were independent and that there was a 50% chance of getting either gender. The significance level was 0.05. A filled-in box signifies that that property in that syllabus was statistically significant, with the color indicating the direction of significance. The blank boxes and the dotted boxes were not statistically significant. A blank box signifies equal distribution.**

| | Statistically significantly more women | More women | Equal | More men | Statisically significantly more men | Total |
|---|---|---|---|---|---|---|
| Pronouns | 2 | 2 | 1 | 12 | 9 | 26 |
| Names | 1 | 1 | 1 | 11 | 12 | 26 |
| Images | 0 | 6 | 6 | 14 | 0 | 26 |
| Audio/video | 0 | 3 | 4 | 12 | 7 | 26 |
| PRM gender | 0 | 8 | 6 | 10 | 2 | 26 |
| **Total** | **3** | **20** | **18** | **59** | **30** | |

**Table 8: Summation of the statistical significance and majorities in Figure 15.**

gender distribution in this category. This finding shows how even in a male-dominated field, it is possible to achieve somewhat equal representation in images. Given the unevenness of the computing industry mentioned above, we can assume that this distribution has been achieved by active choice. Context-sensitive analysis might, again, provide valuable insights.

## 5.4 Observation 4: Nearly 2/3 of all potential role models in our analyzed Danish K-12 Informatics teaching materials corpus are men.

The findings from the role model analysis highlight a noticeable shortcoming in the representation of potential female role models within the teaching material with a 36% women to 64% men ratio. The impact is arguably more significant for potential role models, as substantial research has proved role models to play an important role for novices and women entering IT, e.g., [61, 66, 74, 77, 83]. Given this substantial potential impact, the inclusion of role models stands out as a straightforward area for improvement.

## 5.5 Observation 5: Despite variation between syllabi, women are underrepresented in a majority of individual Danish K-12 Informatics syllabi.

The findings from analyzing the total corpus are reflected in the analyses of the individual syllabi. More syllabi exhibit a majority of men in all properties of the gender representation analysis as well as the role model analysis. This observation indicates that the findings from the total corpus analysis are not created by a few outliers, but rather a coherent picture of the general state of affairs.

## 5.6 Implications for students.

The IT and computing industry is highly male-dominated and this is reflected in the teaching materials in the Danish high school subject Informatics. We see an underrepresentation of women in almost all of the analyses. In the best case, the theoretical content and current examples and assignments (which generally appear relatable to high school students) are enough to spark more students' (including women's) interest in pursuing a career in IT. In the worst case, the lack of female representation in the materials is enough for female students to feel excluded from the computing field before they have even entered it. The truth probably lies somewhere in the middle. With the industry's desperate need for more IT professionals and the stark underrepresentation of women in the field, we argue that changes need to be made to the current available teaching materials.

## 5.7 Implications for teachers.

As digitalization continues, there is an increased need for students to be digitally empowered. They should be able to know digital technology's impact on life and society, its functionality, and how to create it themselves. We believe Informatics-like school subjects will only increase in relevance in public education within the next few years. The teaching materials for these classes should be inclusive of everyone to properly reflect the democratic nature of the subject itself. That means that publishers and teachers need to take an active role in breaking with the idea of computing as a "boys' club", and assume a new course towards fairly representing women in IT.

# 6 THREATS TO VALIDITY

## 6.1 Construct validity

***Observer bias.*** It is essential to recognize our personal involvement in the problem domain. The lead authors are female students in Computer Science and have encountered many of the experiences described in previous literature firsthand. For example, having insufficient early exposure to computing and experiencing a lack of role models from the industry. Additionally, they have been involved in various projects to improve the gender imbalance in computing education over the last six years. With such direct involvement, our objectivity could be questioned. However, the analysis is firmly grounded in the literature and primarily measures objective factors. Several possible criteria, such as the presence of *stereotypes*, were excluded from analysis for the reason of reducing subjective evaluation from the analyses.

***Unit of analysis.*** There is an impact of how "teaching materials" — our unit of analysis — were distinguished, particularly regarding the division of textbooks into sub-sub-chapters. This granularity could influence our results, as the outcomes might have differed if textbooks had been divided differently. The unit of analysis was defined by the syllabi, which often referenced materials at the sub-sub-chapter level. Relying solely on sub-chapter divisions would have been inaccurate, as not all sub-sub-chapters from a sub-chapter were used in syllabi.

***The assessment of people's gender.*** Given the focus of this study on gender representation within teaching materials, an accurate estimation of individuals' gender is key. As a person's gender identity is rarely directly stated, the recording of genders has relied on our assessment. The assessment is based on "traditional" gender presentations of women and men, encompassing visual cues such as hairstyles and clothing, as well as voices and use of pronouns. We do acknowledge that of all the people in the 616 teaching materials, mistakes might have been made, where the assessment has not corresponded to a specific person's gender identity. However, we do not have any reason to believe that errors in gender assessment would be skewed towards one gender over the other.

***Definition of role models.*** Our definition of "potential role model" was informed by existing literature, which suggests that role models can come from various sources and are not one type of person. Opting for a broad definition allows for a more objective analysis, avoiding assumptions about individuals' suitability as role models. Regardless of whether a person would practically be perceived as a role model, the group of "potential role models" can still tell us a lot about the people who represent the computing field in the teaching material.

***The assessment of people's age.*** We recognize the challenge of accurately assessing individuals' ages. While we can't know the exact ages, we made educated guesses and intentionally used broad age categories. Although mistakes may have occurred in our categorization, they likely balanced out in both directions. Additionally, age was not a central aspect in any of our analyses, but rather a supplementary descriptor in our analysis of potential role models.

***Corpus building.*** As explained, listed teaching materials were excluded for various reasons. Of the 123 excluded materials, 88 were due to inaccessibility (broken links, unspecified, and no access), leaving only 35 purposely excluded. It is hard to say if these could have interfered with the results, but since they represent such a small subset of the total number of teaching materials, we find it unlikely.

## 6.2 Internal validity

***Are the collected syllabi representative of HHX?.*** As described, we received syllabi from the school year 2022/2023 from 25% of HHX schools in Denmark. While the exact proportion of HHX Informatics classes covered by our study plans to some degree remains uncertain, we assume that our dataset is reasonably comprehensive. As 41% of the teaching materials were referenced by two or more syllabi, we expect substantial overlap with the remaining HHX Informatics study plans in Denmark.

## 6.3 External validity

***Beyond HHX?.*** As previously mentioned, the official academic objectives in Informatics for the four different types of Danish high schools (STX, HF, HTX, and HHX) appear very similar. While there may be minor variations in curriculum and focus between the high school programs, much of the material examined was not specific to HHX. Only one of the textbooks was specifically made for business Informatics, all others were created for all types of high schools. Consequently, we do not anticipate these differences to significantly impact the extrapolation of our findings to other types of high schools.

***Beyond high school?*** While the teaching materials undoubtedly differ between high school and other K-12 grades, the underlying arguments for gender inclusion remain. Considering the importance of early exposure and how stereotypes are developed early on [11], we might expect an even bigger impact of gender-inclusive teaching in primary schools.

***Beyond binary gender?*** This study adopts a binary gender framework, aligning with existing discourse in the literature. Consequently, our findings are limited to this binary paradigm and do not encompass gender identities beyond it. Paradoxically, this study now contributes to the literature speaking of "gender inclusion" in computing without actually including all genders. We speculate that students who identify outside the gender binary may even be in more need of representation, especially considering that we saw no examples of explicit non-binary representation in the teaching materials. One can only hope that by including a wider range of people in teaching materials, people of all genders will feel more welcomed by the IT field.

***Beyond Denmark?*** This study was a case study, and we can not know if the results generalize beyond Denmark. We believe it is likely that we would see similar results in an equivalent study conducted elsewhere, since Denmark, with a language spoken by less than 6M people, is not large enough to produce a large amount of teaching content independently and will lean on content produced in other countries. Furthermore, although Denmark ranks high on the European Gender Equality Index [1], the ICT gender equality paradox has demonstrated that a disparity in female representation in IT is widely common in Denmark as well as its neighboring countries [23].

## 7 CONCLUSIONS

Our study aimed to answer the following two research questions:

> **RQ1 (Gender representation):** To what extent do Danish K-12 Informatics teaching materials portray equal *gender representation*?

> **RQ2 (Role models):** To what extent do Danish K-12 Informatics teaching materials portray equal gender representation of *role models*?

We found a statistically significant underrepresentation of women in terms of pronouns, names, and presence in audio/video material, but not in images. Women are also underrepresented in terms of potential role models, especially in materials from Danish textbook publishers. Despite some variation between individual syllabi, we see similar patterns between the total corpus and individual syllabi.

The combined findings lead us to the following recommendations, specifically targeted textbook authors and publishers since textbooks, in general, have the largest potential for improvement and have the widest reach of any of the examined formats:

> **Recommendation 1 (Gender representation):** Strive for a more *balanced gender representation* in computing teaching materials; in particular, consider who gets mentioned and in which context.

> **Recommendation 2 (Role models):** Strive for *diversity of role models* in computing teaching materials with varying genders, backgrounds, roles, and interests. It is crucial to confront the implicit assumption about who potentially "belongs" in the IT industry, as the unspoken default often tends to be male.

Adhering to these recommendations could potentially shift some of the burden of responsibility from teachers to textbook authors and publishers. This shift would integrate gender-inclusive teaching materials as an inherent component of Informatics education in the (Danish) K-12 school system, reducing the burden of teachers to ensure gender inclusivity. If we succeed in making gender inclusion an inherent part of all computing education in K-12, it could potentially contribute to more women pursuing a career in the IT field, eventually closing (or diminishing) the gender gap in IT.

## ACKNOWLEDGMENTS

This research was supported by DIREC (Digital Research Centre Denmark) Educational Project P31 on *"Initiatives to improve recruitment and retention of IT students."*

## REFERENCES

[1] [n. d.]. Gender Equality Index | 2023 | DK | European Institute for Gender Equality — eige.europa.eu. https://eige.europa.eu/gender-equality-index/2023/country/DK. [Accessed 25-03-2024].
[2] Karen Aguar, Hamid R Arabnia, Juan B Gutierrez, Walter D Potter, and Thiab R Taha. 2016. Making CS inclusive: An overview of efforts to expand and diversify CS education. In *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, 321–326.
[3] Asli Y. Akbulut. 2016. Majoring in Information Systems: An Examination of Role Model Influence. *Journal of Educational Computing Research* 54, 5 (2016), 660–679. https://doi.org/10.1177/0735633115627825
[4] Catherine T Amelink and Elizabeth G Creamer. 2010. Gender differences in elements of the undergraduate experience that influence satisfaction with the engineering major and the intent to pursue engineering as a career. *Journal of Engineering Education* 99, 1 (2010), 81–92.
[5] Mohadeseh Amini and Parviz Birjandi. 2012. Gender Bias in the Iranian High School EFL Textbooks. *English Language Teaching* 5, 2 (2012), 134–147.
[6] Linda Nørgaard Andersen, Gry Jexen, and Rikke Lie Halberg. 2024. Vores historiebøger rummer en katastrofal skæv fordeling mellem mænd og kvinder. *Politiken* (02 Feb 2024).
[7] L. H Andersen (ISS) and M. K. Andersen (proacteur). 2016. Mangfoldig ledelse giver højere indtjening. http://www.publications.issworld.com/ISS/ISSDanmark/isspresentation/2016/Mangfoldigledelsegiverhjereindtjening/ Available at: http://www.publications.issworld.com/ISS/ISSDanmark/isspresentation/2016/Mangfoldigledelsegiverhjereindtjening/.
[8] Shaki Asgari, Nilanjana Dasgupta, and Jane G Stout. 2012. When do counterstereotypic ingroup members inspire versus deflate? The effect of successful professional women on young women's leadership self-concept. *Personality and Social Psychology Bulletin* 38, 3 (2012), 370–383.
[9] Amanda Barton and Lydia Namatende Sakwa. 2012. The representation of gender in English textbooks in Uganda. *Pedagogy, Culture & Society* 20, 2 (2012), 173–190.
[10] Eric P Bettinger and Bridget Terry Long. 2005. Do faculty serve as role models? The impact of instructor gender on female students. *American Economic Review* 95, 2 (2005), 152–157.
[11] Sylvia Beyer. 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 (2014), 153–192.
[12] Sylvia Beyer and Susan Haller. 2006. Gender differences and intragender differences in computer science students: Are female CS majors more similar to male CS majors or female nonmajors? *Journal of Women and Minorities in Science and Engineering* 12, 4 (2006).
[13] Silvina Bongiovanni. 2014. "No se preocupe la señora marquesa": A study of gender bias in example sentences in the RAE grammar textbook. *IULC Working Papers* 14, 1 (2014).
[14] Jilana S Boston and Andrei Cimpian. 2018. How do we encourage gifted girls to pursue and succeed in science and engineering? *Gifted Child Today* 41, 4 (2018), 196–207.
[15] Olle Bälter, Viggo Kann, Chantal Mutimukwe, and Hans Malmström. 2023. English-medium instruction and impact on academic performance: a randomized control study. *Applied Linguistics Review* (2023). https://doi.org/doi:10.1515/applirev-2022-0093
[16] Børne- og Undervisningsministeriet. 2023. Vejledning til Informatik C, hhx, htx, stx, hf. https://www.uvm.dk/-/media/filer/uvm/udd/gym/pdf23/aug/vejledninger/230822-vejledning-til-informatik-c--hhx--htx--stx--hf.pdf Available at: https://www.uvm.dk/-/media/filer/uvm/udd/gym/pdf23/aug/vejledninger/230822-vejledning-til-informatik-c--hhx--htx--stx--hf.pdf.
[17] Bruce Carrington, Peter Tymms, and Christine Merrell. 2008. Role models, school improvement and the 'gender gap'—do men bring out the best in boys and women the best in girls? *British Educational Research Journal* 34, 3 (2008), 315–327.
[18] SRBH Chaturvedi and RC Shweta. 2015. Evaluation of inter-rater agreement and inter-rater reliability for observational data: an overview of concepts and methods. *Journal of the Indian Academy of Applied Psychology* 41, 3 (2015), 20–27.
[19] Juanjuan Chen, Minhong Wang, Paul A Kirschner, and Chin-Chung Tsai. 2018. The role of collaboration, computer use, learning environments, and supporting strategies in CSCL: A meta-analysis. *Review of Educational Research* 88, 6 (2018), 799–843.
[20] Sapna Cheryan, John Oliver Siy, Marissa Vichayapai, Benjamin J Drury, and Saenam Kim. 2011. Do female and male role models who embody STEM stereotypes hinder women's anticipated success in STEM? *Social psychological and personality science* 2, 6 (2011), 656–664.
[21] Ingrid Maria Christensen, Melissa Høegh Marcher, Paweł Grabarczyk, Therese Graversen, and Claus Brabrand. 2021. Computing educational activities involving people rather than things appeal more to women (Recruitment Perspective). In *Proceedings of the 17th ACM Conference on International Computing Education Research*. 127–144.

[22] Code.org. 2013. What Most Schools Don't Teach. https://www.youtube.com/watch?v=nKIu9yen5nc Available at: https://www.youtube.com/watch?v=nKIu9yen5nc.

[23] Hilde G Corneliussen. 2021. Unpacking the Nordic gender equality paradox in ICT research and innovation. *Feminist Encounters: A Journal of Critical Studies in Culture and Politics* 5, 2 (2021), 25.

[24] Danmarks Lærerforening. 2022. Projekt om kønsopmærksomhed i folkeskolen. https://www.dlf.org/dlf-mener/koensopmaerksomhed/projekt-om-koensopmaerksomhed-i-folkeskolen Available at: https://www.dlf.org/dlf-mener/koensopmaerksomhed/projekt-om-koensopmaerksomhed-i-folkeskolen.

[25] Karen Sofie Egebo. 2020. Sociolog: I Danmark gør vi stort set ingenting for at komme kønsstereotyper til livs. *Kristeligt Dagblad* (14 Jan 2020).

[26] European Commission. 2022. Employed ICT specialists by sex. https://ec.europa.eu/eurostat/databrowser/view/isoc_sks_itsps/default/table?lang=en Available at: https://ec.europa.eu/eurostat/databrowser/view/isoc_sks_itsps/default/table?lang=en.

[27] European Commission. 2023. Employment and Social Developments in Europe 2023.

[28] European Commission. 2023. EU companies face difficulties in hiring ICT experts. https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20230712-1 Available at: https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20230712-1.

[29] Heidi Fencl and Karen R Scheel. 2006. Making sense of retention: An examination of undergraduate women's participation in physics courses. *Removing barriers: Women in academic science, technology, engineering, and mathematics* (2006), 287–302.

[30] GL. 2023. Ligestilling mellem kønnene i gymnasieuddannelserne. https://www.gl.org/detmenergl/Sider/Ligestillingmellem-koennene-i-gymnasieuddannelserne.aspx Available at: https://www.gl.org/detmenergl/Sider/Ligestillingmellem-koennene-i-gymnasieuddannelserne.aspx.

[31] Joanna Goode. 2008. Increasing Diversity in K-12 computer science: Strategies from the field. In *Proceedings of the 39th SIGCSE technical symposium on Computer science education*. 362–366.

[32] Pawel Grabarczyk, Alma Freiesleben, Amanda Bastrup, and Claus Brabrand. 2022. Computing Educational Programmes with more Women are more about People & less about Things. In *Proceedings of the 27th ACM Conference on Innovation and Technology in Computer Science Education Vol. 1*. 172–178.

[33] Virginia Grande, Anne-Kathrin Peters, Mats Daniels, and Matti Tedre. 2018. "Participating under the influence": How role models affect the computing discipline, profession, and student population. In *2018 IEEE frontiers in education conference (FIE)*. IEEE, 1–9.

[34] Carmen Nayeli Guzman, Anne Xu, and Adalbert Gerald Soosai Raj. 2021. Experiences of Non-Native English Speakers Learning Computer Science in a US University. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*. 633–639.

[35] Judith R Halasz and Peter Kaufman. 2008. Sociology as pedagogy: How ideas from the discipline can inform teaching and learning. *Teaching Sociology* 36, 4 (2008), 301–317.

[36] Hedy. 2024. *Textual programming for the classroom*. https://www.hedycode.com/ Available at: https://www.hedycode.com/.

[37] Cedric Herring. 2009. Does diversity pay?: Race, gender, and the business case for diversity. *American sociological review* 74, 2 (2009), 208–224.

[38] https://plus.google.com/+UNESCO. [n. d.]. Explore the ICT Gender Equality Paradox — en.unesco.org. https://en.unesco.org/EQUALS/ICT-GE-paradox. [Accessed 26-03-2024].

[39] Janet S Hyde. 1984. Children's understanding of sexist language. *Developmental psychology* 20, 4 (1984), 697.

[40] Intel and Dalberg. 2016. Decoding Diversity: The Financial And Economic Returns to Diversity in Tech. https://dalberg.com/wp-content/uploads/2016/06/Diversity-report.pdf Available at: https://dalberg.com/wp-content/uploads/2016/06/Diversity-report.pdf.

[41] IRIS Group on behalf of IDA. 2021. Mismatch på arbejdsmarkedet for it-uddannede i 2030.

[42] IT Branchen. 2023. *Beskæftigelse i it-branchen*. https://itb.dk/tema/branchen-i-tal/beskaeftigelse/ Available at: https://itb.dk/tema/it-branchens-talunivers/branchen-i-tal/branchen-i-tal-omsaetning-2-2/.

[43] IT Branchen. 2023. *Løn i it-branchen*. Available at: https://itb.dk/tema/it-branchens-talunivers/branchen-i-tal/loen/.

[44] KEA. [n. d.]. Bliv datamatiker på KEA — youtube.com. https://www.youtube.com/watch?v=5YViIMjjQrs&t=11s. [Accessed 27-03-2024].

[45] Khan Academy. 2022. How to keep your online accounts secure. https://www.youtube.com/watch?v=ukkyJbTSZVQ&t=69s Available at: https://www.youtube.com/watch?v=ukkyJbTSZVQ&t=69s.

[46] Kristin Klopfenstein. 2005. Beyond test scores: The impact of Black teacher role models on rigorous math taking. *Contemporary Economic Policy* 23, 3 (2005), 416–428.

[47] Theodora Koulouri, Stanislao Lauria, and Robert D Macredie. 2014. Teaching introductory programming: A quantitative evaluation of different approaches. *ACM Transactions on Computing Education (TOCE)* 14, 4 (2014), 1–28.

[48] M. Krag-Svendsen (Business Danmark). 2023. *Er du i den bedst betalte branche?* https://businessdanmark.dk/nyheder/artikler/er-du-i-den-bedst-betalte-branche Available at: https://businessdanmark.dk/nyheder/artikler/er-du-i-den-bedst-betalte-branche.

[49] Kvinfo. 2024. Køn og skole. https://kvinfo.dk/koen-i-skolen/ Available at: https://kvinfo.dk/koen-i-skolen/.

[50] Jackie FK Lee. 2014. Gender representation in Hong Kong primary school ELT textbooks–a comparative study. *Gender and Education* 26, 4 (2014), 356–376.

[51] Jackie FK Lee. 2018. Gender representation in Japanese EFL textbooks–a corpus study. *Gender and Education* 30, 3 (2018), 379–395.

[52] Jackie FK Lee and Peter Collins. 2008. Gender voices in Hong Kong English textbooks—Some past and current practices. *Sex Roles* 59 (2008), 127–137.

[53] Jackie FK Lee and Peter Collins. 2009. Australian English-language textbooks: the gender issues. *Gender and education* 21, 4 (2009), 353–370.

[54] Jackie FK Lee and Peter Collins. 2010. Construction of gender: A comparison of Australian and Hong Kong English language textbooks. *Journal of Gender Studies* 19, 2 (2010), 121–137.

[55] Jackie FK Lee and Vahid Mahmoudi-Gahrouei. 2020. Gender representation in instructional materials: A study of Iranian English language textbooks and teachers' voices. *Sexuality & Culture* 24, 4 (2020), 1107–1127.

[56] Richard Lippa. 1998. Gender-related individual differences and the structure of vocational interests: The importance of the people-things dimension. *Journal of Personality and Social Psychology* 74, 4 (1998). https://doi.org/10.1037/0022-3514.74.4.996

[57] Richard A Lippa. 2010. Gender differences in personality and interests: When, where, and why? *Social and personality psychology compass* 4, 11 (2010), 1098–1110.

[58] Penelope Lockwood. 2006. "Someone like me can be successful": Do college students need same-gender role models? *Psychology of women quarterly* 30, 1 (2006), 36–46.

[59] Elizabeth Mannix and Margaret A Neale. 2005. What differences make a difference? The promise and reality of diverse teams in organizations. *Psychological science in the public interest* 6, 2 (2005), 31–55.

[60] Melissa Høegh Marcher, Ingrid Maria Christensen, Paweł Grabarczyk, Therese Graversen, and Claus Brabrand. 2021. Computing educational activities involving people rather than things appeal more to women (cs1 appeal perspective). In *Proceedings of the 17th ACM Conference on International Computing Education Research*. 145–156.

[61] Microsoft Corporation. 2017. Why Europe's Girls Aren't Studying STEM.

[62] Jakob Horn Moller and Lasse Pallesen. 2017. *Computerspil*. Systime, Aarhus C.

[63] Megan M Parise. 2021. Gender, sex, and heteronormativity in high school statistics textbooks. *Mathematics Education Research Journal* 33, 4 (2021), 757–785.

[64] Caroline Criado Perez. 2019. *Invisible women: Data bias in a world designed for men*. Abrams.

[65] Jennifer L Prewitt-Freilino, T Andrew Caswell, and Emmi K Laakso. 2012. The gendering of language: A comparison of gender equality in countries with gendered, natural gender, and genderless languages. *Sex roles* 66, 3 (2012), 268–281.

[66] Eric S Roberts, Marina Kassianidou, and Lilly Irani. 2002. Encouraging women in computer science. *ACM SIGCSE Bulletin* 34, 2 (2002), 84–88.

[67] Els Rommes, Geertjan Overbeek, Ron Scholte, Rutger Engels, and Raymond De Kemp. 2007. 'I'm not Interested in Computers': Gender-based occupational choices of adolescents. *Information, Community and Society* 10, 3 (2007), 299–319.

[68] Raúl Ruiz-Cecilia, Juan Ramón Guijarro-Ojeda, and Carmen Marín-Macías. 2020. Analysis of heteronormativity and gender roles in EFL textbooks. *Sustainability* 13, 1 (2020), 220.

[69] Sandra Schulz, Sarah Berndt, and Anja Hawlitschek. 2023. Exploring students' and lecturers' views on collaboration and cooperation in computer science courses-a qualitative analysis. *Computer Science Education* 33, 3 (2023), 318–341.

[70] Scratch. 2024. *Create stories, games, and animations. Share with others around the world*. https://scratch.mit.edu/ Available at: https://scratch.mit.edu/.

[71] Jørgen Sjaastad. 2012. Sources of Inspiration: The role of significant persons in young people's choice of science in higher education. *International Journal of Science Education* 34, 10 (2012), 1615–1636.

[72] R Patrick Solomon. 1997. Race, role modelling, and representation in teacher education and teaching. *Canadian Journal of Education/Revue canadienne de l'éducation* (1997), 395–410.

[73] Steven E Stemler. 2019. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research, and Evaluation* 9, 1 (2019), 4.

[74] Jane G Stout, Nilanjana Dasgupta, Matthew Hunsinger, and Melissa A McManus. 2011. STEMing the tide: using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of personality and social psychology* 100, 2 (2011), 255.

[75] Rong Su, James Rounds, and Patrick Ian Armstrong. 2009. Men and things, women and people: a meta-analysis of sex differences in interests. *Psychological*

*bulletin* 135, 6 (2009), 859.

[76] Jo Young Switzer. 1990. The impact of generic word choices: An empirical investigation of age-and sex-related differences. *Sex roles* 22 (1990), 69–82.

[77] I Talks, I Edvinsson, and J Birchall. 2019. Programmed Out: The Gender Gap in Technology in Scandinavia. *Plan International Norway: New York, NY, USA* (2019).

[78] The Boston Consulting Group and The Technical University of Munich. 2017. The Mix That Matters: Innovation through Diversity. https://web-assets.bcg.com/img-src/BCG-The-Mix-That-Matters-Apr-2017_tcm9-153209.pdf Available at: https://web-assets.bcg.com/img-src/BCG-The-Mix-That-Matters-Apr-2017_tcm9-153209.pdf.

[79] Uddannelsesguiden. 2024. Merkantil studentereksamen (hhx). https://www.ug.dk/uddannelser/gymnasialeuddannelser/merkantil-studentereksamen-hhx Available at: https://www.ug.dk/uddannelser/gymnasialeuddannelser/merkantil-studentereksamen-hhx.

[80] UX Mastery. 2015. How To Create UX Personas. https://www.youtube.com/watch?v=B23iWg0koi8 Available at: https://www.youtube.com/watch?v=B23iWg0koi8.

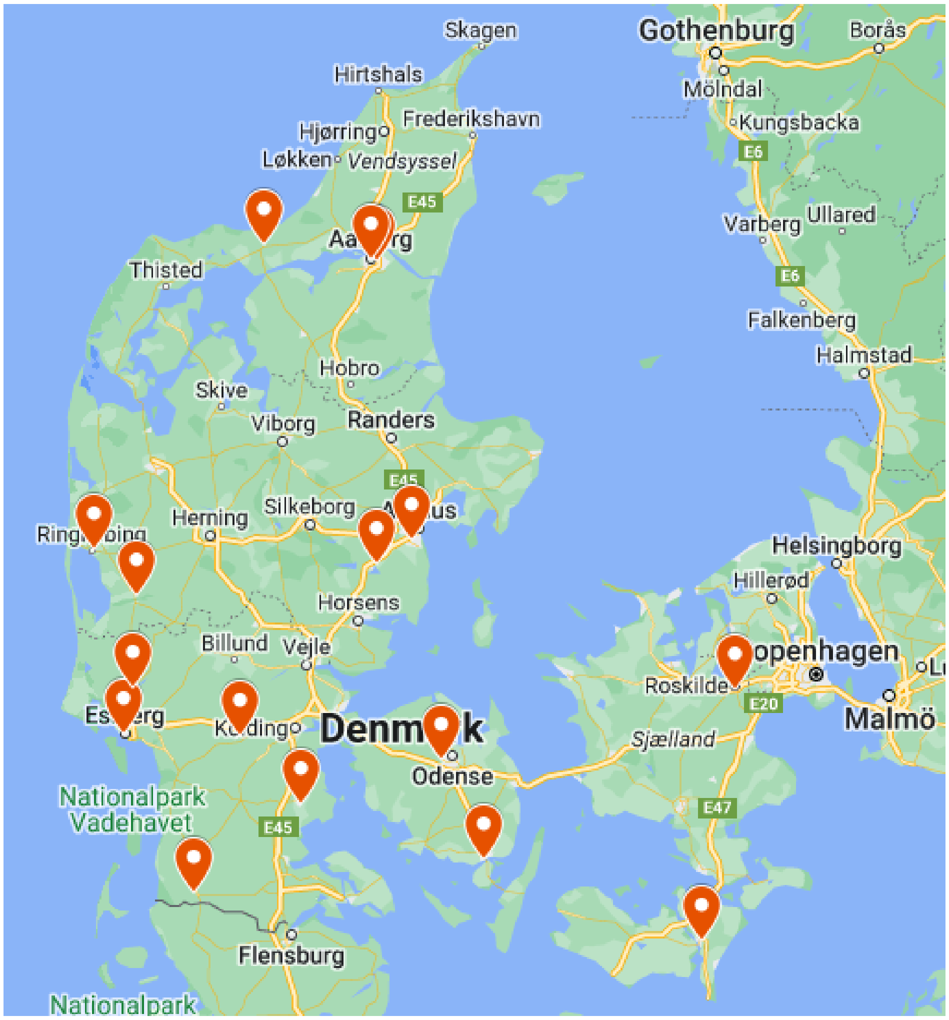
[81] Emiliana Vegas, Michael Hansen, and Brian Fowler. 2023. BUILDING SKILLS FOR LIFE: How to expand and improve computer science education around the world. https://www.brookings.edu/wp-content/uploads/2021/10/Building_skills_for_life.pdf Available at: https://www.brookings.edu/wp-content/uploads/2021/10/Building_skills_for_life.pdf.

[82] Tina Vrieler and Minna Salminen-Karlsson. 2022. A sociocultural perspective on computer science capital and its pedagogical implications in computer science education. *ACM Transactions on Computing Education (TOCE)* 22, 4 (2022), 1–23.

[83] Jennifer Wang and Sepehr Hejazi Moghadam. 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* (Seattle, Washington, USA) *(SIGCSE '17)*. Association for Computing Machinery, 615–620. https://doi.org/10.1145/3017680.3017734

[84] Dexter Yim and Rolian Lee. 2023. A comparative study of gender representation in junior high school EFL textbooks of Hong Kong and Japan. *Journal of Language Teaching* 3, 10 (2023), 21–39.

# A MAP OF HIGH SCHOOLS




**Geographical distribution of the Danish high schools whose materials were analyzed.**