Information Systems and e-Business Management
https://doi.org/10.1007/s10257-024-00696-w

ORIGINAL ARTICLE

Overcome the gender gap: analyzing massive open online courses through the lens of stereotype threat theory

Daniel Stattkus[1] · Lorena Göritz[1] · Katharina-Maria Illgen[1] · Jan Heinrich Beinke[1] · Oliver Thomas[1,2]

Received: 7 August 2023 / Revised: 1 September 2024 / Accepted: 27 November 2024

Abstract

Despite ongoing progress towards gender equality in education, women remain significantly underrepresented in computer science—a field still shaped by stereotypical expectations in our society. This gap is evident in both traditional classroom settings and online learning platforms such as MOOCs, where women face psychological barriers that hinder their learning success. As MOOCs increase accessibility and democratize education, it's particularly important to address the barriers women face in these platforms. By analyzing 338,459 negative reviews from 8,067 IT and software courses offered by the MOOC provider Udemy, we explored the differences in how men and women experience these online learning environments. Our analysis was complemented by ten expert interviews, which helped us develop key propositions to explain these gender-based differences and derive guidelines to overcome them. Our results reveal that men and women criticize similar topics in IT courses, demonstrating that they do not belong to different user groups. However, differences between male and female reviews emerge within each topic. These differences are reflected in different communication styles, demands and areas of emphasis, shaped by gender-specific backgrounds, socialization processes and stereotypes. To overcome these differences, we propose seven guidelines drawing form Stereotype Threat Theory for designing gender-inclusive online courses that focus on inclusive communication and representation, creating supportive learning environments, and implementing high-quality, bias-aware educational practices. Aiming to foster greater participation and success for women in computer science.

Keywords Stereotype threat · Gender · E-learning · Computer science · MOOC

Extended author information available on the last page of the article

Published online: 20 February 2025

1 Introduction

From childhood, societal norms and expectations begin to shape gender identities. Girls are more likely to grow up with dolls and horses, while boys are more likely to grow up with superheroes and ball sports (Blakemore & Centers 2005; Hines et al. 2016). These early experiences not only contribute to individual gender identities but also lay the groundwork for pervasive societal stereotypes that persist today. These stereotypes can negatively affect society; they can constrain individuals within gender roles and discourage them from reaching their true potential. Rather than fostering their inherent talents, individuals might find themselves conforming to society's expectations.

An example is a girl who loves being creative, demonstrates sound logical thinking and problem-solving skills, and is patient and persistent, signifying her potential to excel in programming (Nouri et al. 2020) Despite her natural talents, she chooses not to take a computer science course, primarily because none of her friends have chosen computer science at school, and instead, she takes a course in the social sciences. Similar confrontations with stereotypes are reported by business influencers such as Annahita Esmailzadeh.[1] She recalls her first day at work in a development department when she was looking for her new team. In the corridor, a colleague smiled widely and pointed out to her that the marketing department was on the second floor (Weck 2022). Examples like this illustrate that while girls today actively engage with a wide range of novel information and communication technologies, they are often confronted with stereotypes. As a result, they are still less interested than boys in IT-related careers or college majors (Fedorowicz et al. 2010). The pressure to conform to gender-typical roles and the pervasive threat of stereotypes contribute significantly to this phenomenon (Ashcraft et al. 2016).

Stereotype threat theory (STT) is frequently discussed in the scientific discourse to account for the underrepresentation of women in the information technology (IT) field (Cooper 2006; Takács 2022). Studies dedicated to exploring the impact of gender stereotype threats on performance in computer science tasks have indicated a distinct effect: Women experiencing stereotype threat tend to attribute difficulties internally to their own inability and consequently exhibit a decrease in performance (Cooper 2006; Koch et al. 2008). Recent statistics report a low number (21%) of female recipients of bachelor's degrees in computer science (Datascience@berkeley 2021) and of female learners in computer science courses (Ruipérez-Valiente et al. 2022). This underrepresentation demands rigorous research in computer science to identify more inclusive methodologies for course design and delivery, thereby encouraging increased female participation and success (Ashcraft et al. 2016).

Compared with traditional classrooms, massive open online courses (MOOCs)—through their ease of access—promise greater inclusion and thus the democratization of education (Crues et al. 2018; Kizilcec, Saltarelli, et al. 2017b). In the context of gender-inclusive education, they afford course participants the opportunity to navigate anonymously through the learning environment.

[1] https://www.linkedin.com/in/annahita-esmailzadeh, last accessed on 06.07.2023.

For women in IT and software MOOCs, this anonymity reduces their feeling of being observed while performing computer tasks and their worries about how they will be perceived by others, and consequently, less stereotype threat is triggered (Crues et al. 2018; Lee et al. 2017; Leider & Strobel 2020). Nevertheless, members of identity-threatened groups are still more likely to drop out of IT and software MOOCs than members of groups who do not face such identity threats (Kizilcec, Davis, et al. 2017a; Kizilcec, Saltarelli, et al. 2017b). To explain these inequalities and to provide a holistic solution for the gender-inclusive design of online learning environments, we address the following research questions:

(1) How do men and women differ in terms of challenges when learning computer science skills in MOOCs?
(2) How can the gender-inclusive design of MOOCs contribute to overcoming the gender gap?

We analyzed 338,459 negative reviews from online courses in the "IT & Software" category from the MOOC provider Udemy and derived our results through latent Dirichlet allocation (LDA), a proven method for exploratory analysis of large collections of textual data, followed by a qualitative analysis. We initially focused on textual data to investigate our first research question because, unlike physical barriers in learning environments, psychological barriers such as stereotype threat triggers are primarily influenced by internal psychological processes such as gender socialization. Language can reflect these internal thought processes and thus serve as an object of study (Pennebaker 2011). The decision to focus exclusively on negative reviews stemmed from our aim to uncover and address the underlying psychological barriers that men and women face in these online learning environments. While positive reviews are valuable in highlighting what works well, they tend to provide less insight into the psychological barriers that should be addressed to improve the inclusivity and effectiveness of online IT courses. Following a theoretical sampling approach, positive reviews were therefore excluded, as they would have distorted the problem clusters of the LDA analysis and led to a shift from problem clusters to general clusters about the topics discussed in the reviews. Negative reviews are particularly valuable for identifying problems because they highlight specific unmet expectations or standards (Aithal & Tan, 2021). Moreover, these reviews not only specify issues but also explain why they are problematic, offering a focused and detailed analysis of each fault. This in-depth elucidation often includes causal explanations, enhancing our understanding of the underlying problems and their implications (Ganesan & Zhou, 2016), thereby providing clearer feedback on aspects of the course that failed to satisfy users.

Our methodology, which focused on genuine user feedback from Udemy MOOCs, provided a more pragmatic approach compared with the methodologies employed in previous research, thereby bridging the gap between theory and practice. Advanced techniques for linguistic analysis, such as LDA analysis, allow for meaning to be extracted from valuable comments from course participants. To date, this method has often been used in a commercial context, for example to analyze reviews of hotels (Guo et al. 2017). Our analyzed reviews were clustered into topics containing negative feedback from participants about the course. We identified clusters that addressed the same topic, and based on those clusters, we identified the differences men and women experience in MOOCs. Using expert interviews, we derived propositions to explain these differences and guidelines for the design of gender-inclusive online IT courses. With our guidelines, practitioners will be able to understand which course design options are particularly relevant for the inclusion of women in IT and software courses.

Our paper relies on a binary and social construction perspective of gender that focuses on individuals who self-identify as male or female rather than a non-binary perspective. Thus, we follow previous research regarding the impact of gender on information and communication technology (Stumpf et al. 2020). The main reason for doing so is that the reviews analyzed from Udemy do not allow for any conclusions about gender other than male or female. Therefore, our results leave no room for other identifications, such as non-binary and genderqueer.

Our paper is structured as follows: First, we present the theoretical background in Sects. 2 and 3. Section 4 describes the procedure and results of the LDA analysis we performed on the Udemy reviews. In Sect. 5, we detail the results of our interviews with experts. Finally, in Sects. 6 and 7, we summarize and discuss our main findings and offer an outlook on future research issues.

2 Stereotype threat theory

Stereotypes are beliefs about a certain group of individuals' characteristics, attributes, and behavior (Hilton and Von Hippel 1996). They tend to be inaccurate and do not always apply to this group (Devine 1989). IT-based stereotypes attribute particular IT-related characteristics to a specific user group (Noeltner et al. 2019). For example, IT-based stereotypes apply to older adults (Noeltner et al. 2019), Black people (Cain and Trauth 2013), and women (Cheryan et al. 2013; Clayton et al. 2009; Smith et al. 2005; Yücel and Rızvanoğlu 2019). These groups of individuals are assumed to have difficulties succeeding in computer science. For instance, when a girl or woman wants to acquire skills in computer science, she faces the challenge of making this learning process compatible with her gender identity (Yücel & Rızvanoğlu 2019).

STT aims to explain these types of challenges. According to STT, people in a group feel threatened when handling a task they are assumed to perform poorly on because of a certain stereotype they are aware of in that moment (Steele & Aronson 1995). For example, if a girl or woman performs a difficult computer science task while identifying with her gender identity, she feels threatened to confirm a negative stereotype, which affects her actual performance. This is supported by Koch et al. (2008), who found that women who were reminded that men usually perform better than women in a certain task were affected by stereotype threat, as they attributed the failure to their inability. By contrast, men who were told that women usually perform better than men in a certain task attributed the failure more externally to technical equipment.

STT originates from social psychology and has been employed to explain human behavior during the use of information and communication technology (Appel et al. 2011; Cooper 2006; Koch et al. 2008). Cooper (2006) studied female college students as they completed a computer task wherein the stereotype threat was made either salient or non-salient to them. The results revealed that students who were primed with their identity as a woman immediately before the computer task performed worse than those primed with their student identity. In addition, Kizilcec et al. (2020) found that first impressions influence subsequent behavior. Social identity threat causes a lack of connectedness to the overall group identity and furthers the student's sense of isolation in the class. Therefore, self-doubt must be addressed, as it can continue to hinder a student's success (Leider & Strobel 2020). In computer courses, where stereotypes are salient in specific contexts and women's social identity has been negatively stereotyped, proactive action can be taken by using social cues early in the course to affirm a sense of belonging among members of underrepresented groups (Kizilcec et al. 2020). For example, presenting a male and a female instructor together is an effective strategy for retaining women in such courses (Kizilcec et al. 2020). Furthermore, Cheryan et al. (2011) found that changing the design of an online classroom from one that conveys computer science stereotypes to one that does not significantly increased women's interest in the course and their anticipated success in computer science. Richard and Hoadley (2015) state that supportive communities can improve resilience by mitigating stereotype threat; for instance, environments in which a female-supportive gaming community can foster equitable gaming identification and self-concept can contribute to counteracting the marginalization of underrepresented learners such as women and girls in IT. Brooks et al. (2018) discovered (1) significant evidence that the presence of females in online computer courses induces strong positive effects on overall course activity and discussion posting behavior among female learners and (2) strong evidence of a small negative effect on male learners. All the evidence suggests that subtly personalized interventions in educational environments and computing identity can influence engagement in online learning environments among minoritized learners in computer science, such as women (Clarke et al. 2023), and that course setup can encourage positive group identity in online learning environments where discrimination based on, for example, gender is predominant (Leider and Strobel 2020).

Our research contributes to the discourse on STT by extending it into the domain of online learning, specifically through the lens of MOOCs. By analyzing negative feedback from Udemy courses, we provide insight into the experiences of women in IT education, showcasing how stereotypes hinder their engagement and performance. This study confirms the presence of stereotype threat in digital learning environments and proposes practical guidelines for designing MOOCs that mitigate this threat. By linking the theoretical foundations of STT and its practical implications for online education, our study provides a pathway to more inclusive learning experiences.

3 Gender-inclusive design of MOOCs

Uncertainties that arise when students engage with stereotypical learning content that might threaten their social identity are what research refers to as psychological barriers (Easterbrook and Hadden 2021; Kizilcec and Saltarelli 2019b). Easterbrook and Hadden (2021) characterize these psychological barriers as a sense of threat to social identity and a feeling of incongruence between student identity and educational success. Psychological barriers in the context of online learning include cues in learning environments, such as verbal, visual, or behavioral artifacts, that may unconsciously threaten learners' social identities by signaling the importance of certain social identities, such as majorities, for a given subject (Kizilcec and Saltarelli 2019b). These barriers can leave members of an underrepresented group feeling alienated and discouraged, which affects their persistence and performance within education (Easterbrook et al. 2019) and consequently contributes to educational inequality (Walton and Yeager 2020).

Kizilcec and Saltarelli (2019b) propose a psychologically inclusive design as an evidence-based approach to providing learning environments for all students to reduce uncertainty in the learning process. Within this approach, content and design cues, such as images or text, are strategically placed within the learning environment to avoid threatening students' social identities and reinforcing their insecurities. These small, subtle, and often easy-to-implement design solutions aim to improve the subjective learning experience of students, especially those who feel they do not belong. In this way, their educational experience can be transformed from one of fear and threat to one of safety and confidence (Easterbrook & Hadden 2021).

Previous research has demonstrated that brief psychological interventions can improve the learning process of underrepresented groups, such as women in male-dominated fields (Brooks et al. 2018; Leider and Strobel 2020; Miyake et al. 2010; Kizilcec and Saltarelli 2019a). Given the digital transformation in education, Kizilcec and Saltarelli (2019b) emphasize the need for further research on the design of psychologically inclusive online learning environments. Previous studies have shown that the significant increase in e-learning platforms and the growing interest in learning to program makes it increasingly important to guarantee that today's e-learning environments meet the needs of all members of society (Astleitner and Steinberg 2005; Kelleher et al. 2007). MOOCs are a form of e-learning and make learning content accessible to anyone, anytime, anywhere (John and Meinel 2020). Because of this easy access, in the context of gender inclusivity, MOOCs have been heralded as a vehicle for democratizing education (Crues et al. 2018; Kizilcec, Saltarelli, et al. 2017a, b). In addition, previous research has argued that the anonymous and faceless MOOC learning space often leaves information about students' gender unknown, thereby eliminating any concerns about how they might be perceived by others (Crues et al. 2018; Lee et al. 2017; Leider and Strobel 2020). However, compared to threatened groups, members of identity-threatened groups drop out of IT and software MOOCs more frequently, which might be a result of psychological barriers (Kizilcec, Davis, et al., 2017; Kizilcec, Saltarelli, et al. 2017a, b). To some extent, IT and software MOOCs face similar challenges to traditional classrooms in that women must also make stereotypically male learning content compatible with their gender identity (Lee et al. 2017). However, Brooks et al. (2020) and Kizilcec and Saltarelli (2019b) further emphasize that large-scale online learning environments offer new opportunities to address this challenge to achieve greater inclusion in education. As MOOCs can reach large numbers of learners, a small adjustment to the learning environment immediately impacts the experience of many students worldwide. As a result, changes in online learning environments are faster, less costly, and thus more efficient to implement than physical learning environments. Furthermore, applying artificial intelligence to learning platforms offers the opportunity to address the need for increased inclusion in education. For example, implementing algorithms that recommend tailored learning content or automatically adapt the learning environment to learners' needs enables course instructors to incorporate differentiation into the course successfully. In addition to content and design cues, online course designers can also easily customize the level of interaction in the learning environment to meet learners' needs, for example by allowing them to interact anonymously in the learning environment. Psychologically inclusive design aims to provide concrete recommendations for supporting underrepresented groups.

In recent years, a feminist human–computer interaction (HCI) community has emerged with papers published (e.g., at the Conference on Human Factors in Computing Systems) on how gender-based differences should influence the design of digital environments (Bardzell & Bardzell 2011; Fiesler et al. 2016). One developed and formalized method for considering gender in technology design is the GenderMag method (Burnett et al. 2016). It provides HCI designers with step-by-step instructions and ready-to-use forms to uncover gender barriers in existing systems. In their conceptual review, Stumpf et al. (2020) overview previous research on gender-inclusive HCI design. They emphasize that the first step is to uncover and examine implicit assumptions and stereotypes that are already built into the technology. John and Meinel (2020) provide recommendations, in their case study, specifically to support women engaging in IT MOOCs. Bhargava (2002) developed a checklist to help evaluate educational software to promote girls' interest in IT. Moreover, Göritz et al. (2022) followed an automated approach to evaluate educational resources and developed a tool that automatically analyzes textual gender bias. To ensure that gender is increasingly considered in the design process, we want to contribute to gender-inclusive design research on MOOCs.

We contribute to the discourse on the gender-inclusive design of MOOCs through an analysis of negative course reviews, and our research sheds light on the psychological barriers that disproportionately affect women in IT MOOCs. This approach allows us to pinpoint where design choices inadvertently reinforce stereotypes, thus allowing us to provide guidelines for effective design solutions for gender inclusivity. Our study therefore contributes to a broader effort to ensure MOOCs serve as a truly democratizing force in education.

4 Latent Dirichlet allocation

In this section, we explore the application of LDA as a methodological approach to uncovering underlying topics within our dataset of online course reviews. Utilizing Udemy's extensive repository of course data, we employed LDA to cluster negative reviews, providing a nuanced understanding of challenges encountered in online IT and software courses. We used the course reviews as archival data to gain initial insight into our researched phenomenon as part of our grounded theory research approach. Grounded theory refers to a set of strategies used to develop a theory through the simultaneous collection and analysis of data and the abductive interplay of induction and deduction (Glaser et al. 1968). As a first step, we conducted the LDA analysis to gain inductive access to theoretical propositions related to our research questions. Our rigorous analysis allowed us to elucidate distinct themes and gender-specific patterns in learners' critiques, offering valuable insights into the landscape of MOOCs and potential areas for improvement.

4.1 Methods

Udemy as a data source. To obtain the necessary data for the analysis, we requested API access to various large MOOC providers, which Udemy granted. According to their own information, with more than 62 million users and over 210,000 courses, Udemy is among the leading providers of online courses.[2] During the COVID-19 pandemic, Udemy courses were the most searched on Google (Kansal et al. 2021). Udemy provides a wide range of courses and a representative course landscape. We used Udemy as a representative data source to draw conclusions about MOOCs in general.

The data used in the analysis was obtained via the Udemy Affiliate API. Unlike scraping data, using the API minimizes error vulnerability during data extraction. The extracted dataset opened great potential for analyzing stereotype threat in online education. It contained many different course reviews, which allowed for a quantitative analysis. Such a broad dataset is rare in the education domain, as it is subject to privacy regulations on the one hand and disclosure from MOOC operators on the other. We only encountered some limitations in accessing ratings and courses, restricting our analysis to a subset of the top 10,000 courses on Udemy. However, the technical limitations in accessing ratings and courses primarily impacted only the top 47 courses with more than 10,000 ratings, which is negligible and possibly even beneficial, as these courses with more than 10,000 ratings are not overrepresented in the analysis. Despite this constraint, we anticipate high generalizability due to the vast amount of user data, the diversity of the courses analyzed, and Udemy's market position as one of the largest MOOC providers. For the top 10,000 courses in the "IT & Software" category, the course ID and the course title were extracted, each with 10,000 ratings accessible over the ID via the API. The rating data included the reviewer's user ID, name, rating (0.5 to 5 stars), and comments.

[2] https://about.udemy.com/, last accessed on 06.07.2023.

Retrieving additional metadata such as a user's gender, age, or other courses taken would have been interesting for deriving the propositions, but such data was not publicly available. This limitation is primarily due to the data access policy of the Udemy Affiliate API, which does not provide such specific user information. This restriction is in line with broader industry practices around user privacy and data protection, where platforms are increasingly cautious about the types of user-related data they disclose. Despite this limitation, the data available through the Udemy Affiliate API can effectively address our primary research questions.

Data preparation. To enable an analysis of the data, prior data preparation was necessary. The goal of this data preparation was to filter out irrelevant data records, clean up errors, and add information necessary for the analysis, such as the sentiment of the review and the gender of the user. Data preparation began with the removal of all ratings without any accompanying comments from the dataset. Next, the names provided by the user were cleaned. If the user provided a name that did not contain any letters (e.g., only dots), it was declared "Anonym." The dataset had a disadvantage, however: It did not include the gender of the users. Due to the great potential of the data, we nevertheless decided to use it for the following analyses. Since the analyses relied on a division into male and female reviews, the gender had to be predicted in advance.

Due to the size of the dataset, we chose an automatic approach for the analysis. Users who used a designation or abbreviation in their name that could be assigned to gender (e.g., Ms., Mr., or Sir) were declared directly to be the corresponding gender. To minimize potential errors, gender-neutral titles (e.g., Dr.) were excluded from this classification process. We acknowledge that this approach, while efficient, is not without its limitations. The method may introduce some degree of error, but such instances were expected to be minimal and unlikely to significantly affect the overall findings of our study. To minimize errors, we screened the validity of random samples that were predicted, both manually and in comparison with the approach used for the remaining names without any designation or abbreviation. For these names, three approaches summarized by Vasilescu et al. (2012) were considered for further gender detection.

We considered a writing-style-based approach that predicts gender based on the words used in written text. The approach is based on the fact that pronouns such as "I," "you," and "she" are significantly more often used by female authors, whereas male writers use "of" phrases such as "garden of roses" more frequently (Vasilescu et al. 2012). Although this approach achieves precision rates of up to 80%, it was unsuitable for the present dataset. Reviews of online courses are not comparable in writing style and length to fiction, scientific papers, or blog posts, which were used as a foundation for this analysis. A profile picture-based approach, which is often used to detect the gender of users, was further considered. Since the dataset did not contain any profile pictures, it was impossible to make a prediction based on them. Therefore, our analysis followed the approach of predicting gender by name.

The name of a person often allows for an accurate gender prediction. As previously stated, however, this approach did not allow for the identification of non-binary or genderqueer people, which is why our analysis focused on a binary perspective.

Vasilescu et al's. (2012) GenderComputer[3] tool, developed for the analysis of Stack Overflow comments, was used for the analysis. The tool predicts a user's gender based on their name and origin. For example, the name Andrea is more likely to be given to women in Germany, while it is a male name in Italy. Since the user's origin in this dataset was unknown, the GenderComputer would output "gender unknown" for the name Andrea. If name and gender are identical in all countries contained in the dataset of the tool, it would return either "male," "female," or "unisex" for the name. In their 2014 study, Vasilescu et al. (2014) achieved 93% precision in predicting gender. We consider this precision rate sufficient in the context of gender prediction, as the GenderComputer has successfully been used in various other studies (Amendola et al. 2024; Dubois et al. 2020; Jarvis et al. 2023; Vasilescu et al. 2014). To reduce mismatches, all participant reviews labeled as either unknown or unisex were excluded from the analysis. While this approach increased the precision of predictions, it thinned out the dataset – a trade-off that could be taken due to the number of data points available. While we still cannot guarantee that every name was correctly assigned, we are confident that, on average, sufficient precision was achieved.

The GenderComputer tool was used for the remaining names to determine whether a name is more likely to be female, male, unisex, or unknown. Subsequently, duplicates were removed from the dataset. These duplicates mainly occurred when a coherent course was separated into sub-courses. To obtain consistent results, the focus of the analysis was on English reviews only. All non-English reviews were therefore filtered from the dataset.[4] Following the filtering, all characters that were neither Latin letters nor spaces were removed from the reviews. Moreover, all uppercase letters were changed to lowercase. Finally, the sentiment of the individual reviews was determined using the VADER Sentiment analysis tool, which operates through a blend of lexical items and rule-based methods to evaluate sentiment polarity and intensity. VADER employs a sentiment lexicon comprising words, phrases, and emoticons, among other things, each rated for sentiment valence on a scale from − 4 (extremely negative) to + 4 (extremely positive), with 0 being neutral. The sentiment score of a sentence is calculated by breaking down the sentence into tokens and matching each against the lexicon, adjusting the overall sentiment based on the intensity of the identified tokens. Additionally, VADER applies heuristic rules to account for grammatical and syntactical cues that might modify sentiment intensity, such as punctuation emphasis, capitalization, degree modifiers, the contrastive conjunction "but," and negation. The final step involves computing a composite sentiment score, ranging from -1 (most negative) to + 1 (most positive), to classify the sentence as positive, neutral, or negative. This methodology allows for nuanced sentiment analysis, especially in online contexts where expressive language forms (i.e., slang) prevail (Hutto & Gilbert 2014). The sentiment scores from VADER were aggregated over the ratings, with an average sentiment value calculated for each rating category, as shown in Table 1.

[3] https://github.com/tue-mdse/genderComputer, accessed on 06.07.2023.

[4] Using the Python library langdetect: https://pypi.org/project/langdetect/, last accessed on 06.07.2023.

Table 1 Sentiment Scores by Rating

Rating	> 5	> 4.5	> 4	> 3.5	> 3	> 2.5	> 2	> 1.5	> 1
Sentiment	0.355	0.270	0.117	0.031	– 0.069	– 0,112	– 0.151	– 0.167	– 0.149

Based on the resulting sentiment per rating group, only datasets with a rating of fewer than three stars were retained. For those, the sentiment analysis results became, on average, negative and therefore support the creation of problem clusters and do not distort the LDA analysis towards the creation of general clusters about the topics discussed in the reviews. In the last step of data preparation, the dataset was split into two sets. One dataset contained only reviews written by users declared as female, and the other, only reviews written by users declared as male. Therefore, the analysis did not include users with names identified as unisex or with names that could not be assigned. Of the original 2,811,475 reviews from 10,000 courses, 338,459 (12.04%) reviews from 8,067 (80.67%) courses remained after cleaning and filtering. Of the remaining reviews, 63,953 (18.9%) were assigned to females and 274,506 (81.1%) to males. Figure 1 presents an overview of the entire data preparation process.

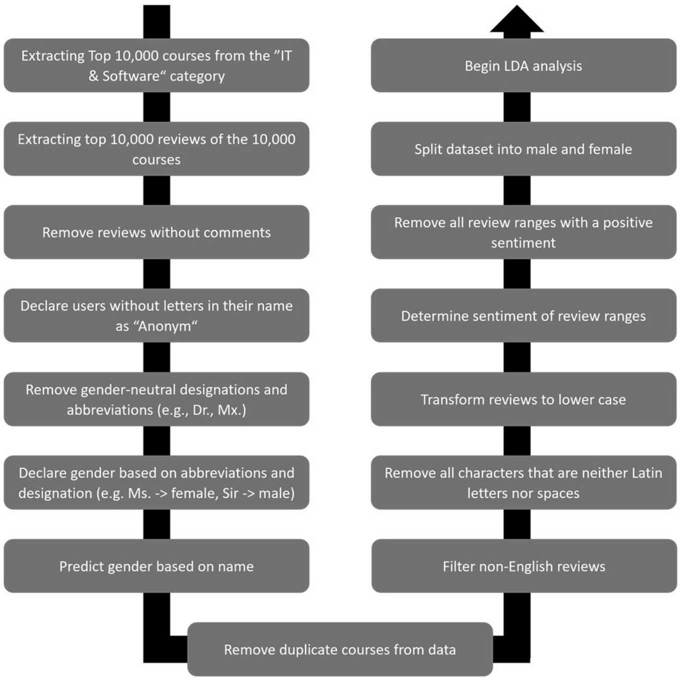

Fig. 1 Data preparation

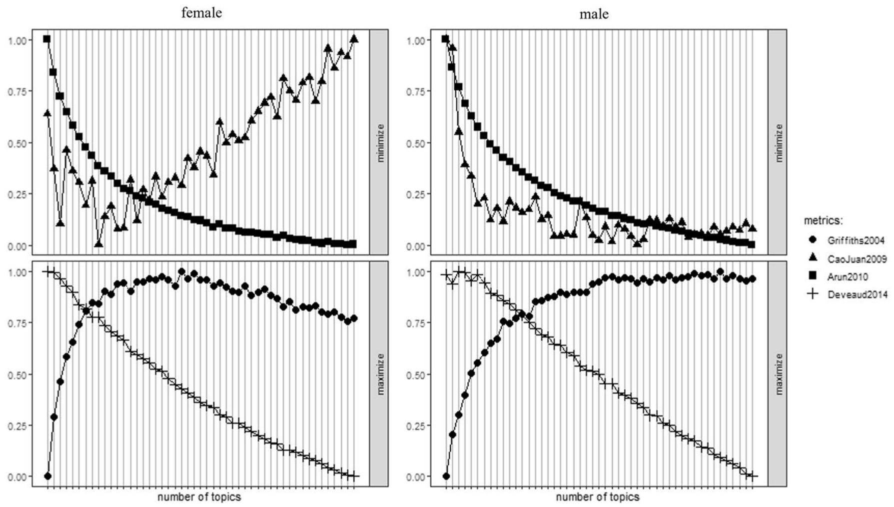

Fig. 2 Model fit metrics

Review clustering. In the next step, we used LDA to cluster the negative reviews into topics. LDA is a widely used method for the exploratory analysis of large collections of textual data. It is a probabilistic, generative statistical model by Blei et al. (2003) that conceptualizes each document as a finite mixture of topics and each topic as an infinite mixture of words. This perspective allows for the effective decomposition of the corpus into latent topics based on word co-occurrence patterns, providing insights into the thematic structure underlying the textual data. By assuming that documents are generated through a specific stochastic process involving Dirichlet-distributed topic mixtures, LDA facilitates the uncovering of hidden thematic layers within the data (Blei et al. 2003). In our study, these topics address challenges in learning computer science skills in MOOCs. We calculated different metrics for several LDA models to determine the best number of topics for the model.[5] The results of these metrics are illustrated in Fig. 2, where the *x*-axis represents all models with a topic number from 2 to 50.

Low values according to the metrics of Arun et al. (2010) and Deveaud et al. (2014) and high values according to the metrics of Griffiths and Steyvers (2004) and Cao et al. (2009) indicate a “good model” that contains distinguishable topics. According to the metrics, the best model solutions ranged between 10 and 16 topics for all female reviews and 19 and 28 topics for all male reviews, respectively. For all models within this range, the LDA was performed to investigate which number of topics was the best solution. The decision to segment the data by gender before applying LDA was driven by the significant disparity in data volume — 81.1% of the 338,459 reviews were written by male participants. This imbalance skewed the overall data distribution and impacted the granularity of topic detection within each

[5] Using the R package ldatuning: https://cran.r-project.org/web/packages/ldatuning/index.html, last accessed on 06.07.2023.

Table 2 Topic Content

Topic Field	Topic Themes	Topic Number	Topic Description	Accuracy Score
Course Framework	Course Presentation	1female	The instructor is criticized for their language, primarily a strong and poorly understandable accent. Non-existent or incorrect subtitles, grammatical errors, and poor speaking skills make comprehension difficult.	98%
		2female	Audio quality is criticized for distracting noise as well as inconsistent or too low volume, while video quality is criticized for blurry images, grammatical errors, or poor visual support.	94%
		3female	The instructor is criticized for poor presentation skills, both in terms of a boring speaking style, partly because they only read the script or speak too quickly, and a poor use of presentation tools, such as PowerPoint.	86%
	Value of the Course	4female	The cost–benefit ratio is criticized. The course participants are so disappointed with the quality and/or quantity of the course content that they describe it as a waste of both time and money and, in some cases, demand their money back.	82%
Course Content		5female	The course is criticized for sometimes not being worth the money, as there are already several free alternatives on the internet.	50%
	Expectations	6female	The course is criticized for not meeting users' quality-related expectations in terms of actuality, test preparation, lecture style, and content.	88%
		7female	The course is criticized for not meeting the course expectations. The level of difficulty is not aligned with the user's expectation, which is based on the course title/description. The course is either too easy or too difficult, such that specific prior knowledge is necessary. Furthermore, expectations regarding the content are not met. The criticisms are mainly related to completeness, quality, and the instructor's teaching style.	80%
	Explanations	8female	The instructor is criticized for explanations that are missing, lack sufficient depth, or are too detailed in the wrong places and for not answering questions.	64%
		9female	The instructor is criticized for their explanations. They explain poorly, and participants consequently often feel unprepared for the "real world."	60%
		10female	The instructor is criticized for explaining coding poorly.	58%
		11female	The course is criticized for using outdated materials and providing insufficient explanations and instructions on installing software.	70%
Exam Preparation	Exam Preparation	12female	The tests are criticized for inadequately preparing students for real-world certification exams, predominantly in cloud computing.	88%
		13female	The practice tests are criticized. They contain grammar and spelling errors; typos; and repetitive, incorrect, or misleading questions and answers.	100%
Course Framework	Course Presentation	1male	The instructor is criticized for their language, primarily a strong and poorly understandable accent. Non-existent or incorrect subtitles, grammatical errors, and poor speaking skills make comprehension difficult.	100%
		2male	The audio quality of the course videos is criticized due to distracting noise, inconsistent or too low volume, unsuitable background music, and a poor microphone.	98%
		3male	Video quality is criticized, for instance, if there are too many very short videos or if the videos are unnecessarily long due to irrelevant content and repetition. The visual presentation is affected by frozen or blurred images and poor video editing.	82%
		4male	The instructor is criticized for poor presentation skills, both in terms of a boring speaking style – because they only read the script – and a lack of visuals, for example in the presentation.	100%
		5male	The instructor is criticized for speaking too fast, or the overall course speed is criticized for being either too fast or too slow and repetitive.	76%
	Value of the Course	6male	The cost–benefit ratio is criticized. The course participants are so disappointed with the quality and/or quantity of the course content that they describe it as a waste of both time and money and, in some cases, demand their money back.	96%
Course Content		7male	The course is criticized for sometimes not being worth the money, as there are already many free alternatives on the internet.	62%
	Expectations	8male	The course is criticized for not meeting the course expectations. The level of difficulty is not aligned with the user's expectation, which is based on the course title/description. The course is either too easy or too difficult, such that specific prior knowledge is necessary. Furthermore, expectations regarding the content are not met.	88%
		9male	The course is criticized for not meeting the course expectations. Users are unsatisfied with the learning material, the completeness of the course content, and the instructor's teaching skills.	90%
		10male	The level of difficulty is criticized for not meeting the user's expectation, which is based on the course title/description. The course is either too easy or too difficult, such that specific prior knowledge is necessary.	84%
	Explanations	11male	The instructor is criticized for explanations that are missing, lack sufficient depth, or are too detailed in the wrong places. Course participants often lack the "why" and visual support to understand the concepts explained.	82%
		12male	The instructor is criticized for teaching with too few useful examples, such that participants often feel unprepared for the "real world."	74%
		13male	The instructor is criticized for making coding mistakes. They either do not explain coding examples correctly or give course participants incorrect coding tasks.	82%
		14male	The instructor is criticized for several aspects that can be summarized as teaching skills.	98%
		15male	The software and hardware configurations are criticized; the course contains errors, including in the tests; learning content is not sufficiently delivered; or necessary information is missing.	100%
	Support	18male	The lack of support regarding questions from course participants and the timeliness of course content are criticized.	82%
Exam Preparation	Exam Preparation	16male	The tests are criticized for inadequately preparing students for real-world certification exams, predominantly in cloud computing.	100%
		17male	The practice tests are criticized. They contain grammar and spelling errors; typos; and repetitive, incorrect, or misleading questions and answers.	100%

subset. We adopted this approach based on findings from a study by Puyol-Antón et al. (2021) that compared several methods to address bias in imbalanced datasets. The study concluded that segmenting data by groups yielded the best balance between accuracy and fairness. Further, by segregating the data, we aimed to prevent male-authored reviews from diluting the distinctiveness of topics more prevalent in female-authored reviews. Consequently, separating the data by gender allowed for a more nuanced understanding of thematic structures that could be obscured by modelling unsegmented data. This approach therefore mitigates the risk of overshadowing unique topics primarily discussed by female reviewers, which might not surface in a mixed-gender analysis due to their relative scarcity. However, segmenting the data led us to identify more distinct topics in the male reviews than in the female reviews. We analyzed typical words for each topic to determine the model in which the topics were most discriminative. Based on these results, we decided to use a model with 14 topics for the female reviews and a model with 19 topics for the male reviews. For the LDA, we used the Gibbs sampling algorithm (Geman & Geman 1984).[6] The LDA output of all female reviews was the weight of each word and each review for 14 topics. The LDA of all male reviews resulted in the weight of each word and each review for 19 topics. To explain the co-occurrence of words in the reviews, we conducted a manual qualitative analysis (Mayring 2010): We worked in a group of five researchers who went through every round of coding and discussed our results in joint workshops after each step. As advised in grounded theory research, the researchers were provided with review information (e.g., the gender of the reviewers) to establish conceptual proximity to the phenomenon under investigation and to enable a deep examination of social processes from an "insider" position (Murphy et al. 2017). While this information helped us contextualize the data, we set the rules of analysis to remain anchored in the description of the content of

[6] Using the R topicmodels: https://cran.r-project.org/web/packages/topicmodels/index.html, last accessed on 06.07.2023.

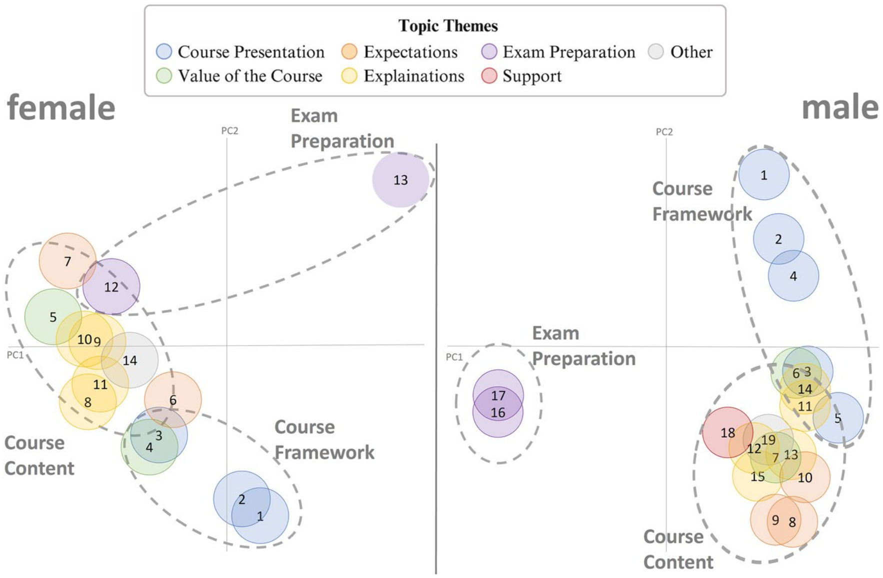

Fig. 3 Intertopic distance map

the reviews themselves. This ensures that the summarization of the topic content, derived from LDA, is grounded in the data rather than subjective and preconceived notions about gender. To identify the themes underpinning each topic, we followed a typical approach and used two criteria for analysis (Baumer et al. 2017). In the first round of coding, we inductively examined the top 25 most probable words per topic, described patterns in these words, discussed them, and assigned initial labels to the topics. In the second round of coding, we sorted the reviews by their concentration on each topic, analyzed the top 50 reviews, and refined coherent and interpretable descriptions of the topics in collaborative workshops. We further clustered the topic descriptions into topic themes and topic fields, as listed in Table 2.

4.2 Results

Table 2 overviews the different parts of our results (topic description, topic theme, and topic field). In addition, each topic is assigned a topic number and an accuracy score indicating what percentage of the 50 highest-weighted reviews match the selected topic description. The intertopic distance map in Fig. 3 illustrates how the overarching topic fields, namely course framework, course content, and exam preparation, emerged. It also depicts the topic themes in different colors, with the same topics for female and male learners marked in the same colors for better understanding of similar topics. The map extracts information from a fitted LDA topic model and visualizes the topics as circles in a two-dimensional plane.[7] The centers of the

[7] Using the R package LDAvis: https://cran.r-project.org/web/packages/LDAvis/index.html, last accessed on 06.07.2023.

circles are determined by calculating the Jensen–Shannon divergence between the topics (Lin 1991). Multidimensional scaling is used to map the distances between topics into two dimensions. In our analysis process, we found that many topics mentioned in the male reviews also appeared in the female reviews. However, because we identified a greater number of distinct topics in the male reviews overall, certain criticisms were separated into distinct topics in the male reviews (e.g., audio and video quality in the topics *2male* and *3male*), while the same criticisms were consolidated into a single topic in the female reviews (e.g., audio and video quality in the topic *2female*). We discussed these differences in our workshops, thoroughly examined the analyzed reviews, and concluded that they still provide a way to compare similar content across genders. To present the results, we describe the female topics one by one, and for each of these topics, we use similar male topics for comparison, thereby allowing for the identification of gender-based differences and similarities regarding the challenges of online IT and software courses. At the end of this section, we summarize our findings by aggregating them into three areas of gender-based differences (Table 3).

Course framework. Topics *1female* and *1male* mainly refer to the challenge of not understanding the instructor's accent, as English was often not the mother tongue. In addition, the strong accent led to incorrectly generated subtitles and, thus, also to greater difficulties in understanding.

Topic *2female* refers to the audio and video quality of the courses. On the one hand, audio quality was criticized for distracting background noise or poorly adjusted volume. On the other hand, video quality was criticized for containing blurry images or incorrect or poor visuals. These topics were also found in the results of the men's analysis. However, the audio and video quality themes appeared separately in two different topics: *2male* and *3male*.

Topic *3female* relates to the instructor's boring presentation skills, partly because the instructor read everything from the script and partly because of poor use of presentation tools such as PowerPoint. Some of the female reviews analyzed also mentioned the instructor speaking too fast. For men, presentation skills and speech rate were mapped into two separate topics: *4male* and *5male*. The qualitative analysis of the results revealed a tendency for the female reviews to offer more suggestions for improvement (e.g., "More graphics and visual aids apart from reading from the PowerPoint slides could be useful"). Regarding the criticized speed of the instructor's speech, female course participants mainly perceived speaking as too fast. Speed was criticized much more extensively in topic *5male* and included criticisms that the instructor spoke too fast and that the overall course speed was either too fast or too slow and repetitive.

Topic *4female* concerns the cost–benefit ratio: Course participants were disappointed with the quality and/or quantity of course content. The courses were often described as a waste of time and money, and in some cases, course participants demanded their money back. Topic *6male* comprises criticisms of the same aspects. However, in comparison to females, males were harsher when demanding their money back. On average, male reviews were more represented by statements such as "I request immediate refund" or direct attacks on the instructor (e.g., "Total waste of money. The guy just rambles on and on because he likes the sound of his voice"),

Table 3 Overview of gender-specific differences in reviews

Area of gender-based difference	Finding	Topic	Topic theme	Topic field
Communication Style and feedback culture	Women are more likely to criticize the didactic approach of teachers	8–11female, 11-15male	Explanations	Course content
	Women make more direct suggestions for improving presentation skills	3female, 4male, 5male	Course presentation	Course framework
	Women give more concrete examples of free alternative courses	5female, 7male	Value of the course	Course framework
	Men list more mistakes that the teacher makes in the explanations	10female, 13male	Explanations	Course content
	The reviews by male participants are between 21 and 166% longer than those by female participants (especially in clusters with a technical focus)	All, especially, 10female, 13male	All	All
	Men are harsher and more assertive when asking for their money back	6male, 4female	Value of the course	Course framework
Asking for support	Female participants do not criticize the instructor for not being available for questions or not regularly updating the course content	Absent	Support	Course content
	Male participants criticize the instructor for neglecting the course, being less available for queries, and not updating the course regularly	18male	Support	Course content

Table 3 (continued)

Area of gender-based difference	Finding	Topic	Topic theme	Topic field
Exam preparation	Female participants place a higher emphasis on understanding explanations as a part of their exam preparation, indicated by the proximity of topics related to exam preparation and explanations	12female	Exam preparation	Exam preparation
	Male participants prioritize other aspects of exam preparation, as indicated by the greater distance between these topics in their feedback	16male	Exam preparation	Exam preparation
	Male participants critique the adequacy of course preparation for certification exams	16male	Exam preparation	Exam preparation

while female reviews expressed dissatisfaction without the same harshness, with statements such as “Please refund my amount in bank account” and “Please don’t waste money in this.”

Course content. Topic *5female* concerns criticisms that alternatives on the internet grant participants free access to the course content. Because of these alternatives, participants felt the reviewed course was not worth the money. These criticisms were also found in topic *7male*. However, the qualitative analysis of the topics revealed that concrete examples of alternative courses were explicitly mentioned more in the feedback of female than male course participants.

For both males and females, there were several topics related to the topic theme Expectations (*6female, 7female, 8male, 9male,* and *10male*). The female reviews criticized some aspects that were not directly mentioned in many male reviews, such as the topicality and quality of the content. While male course participants only implicitly mentioned these points, female course participants explicitly mentioned them via criticism of the existing teaching material (e.g., “I find that lots of this content is out of date. This can be subtle like the console appearance has changed, which is pretty inconsequential, but other times there are additional/new options or even the location of items within the console has changed completely”).

The topic theme Explanations is also reflected in multiple topics (*8female, 9female, 10female, 11female, 11male, 12male, 13male, 14male,* and *15male*). These topics address general explanations by teachers and criticisms related to explaining programming or configuring hardware and software. Notably, criticisms from male and female course participants differed. The lack of or poor visual support was criticized more often in male reviews than in female reviews. In addition, male participants more often explicitly mentioned that the lecturer taught with too few or poor examples from practice. Males and females also differed in terms of how they criticized. Female course participants displayed a greater tendency to describe how they would like the programming to be explained to them (e.g., “There needs to be more direction in what is going to be coded. For example, there should have been a Class Diagram before creating the classes out of the blue”), while male participants listed more errors (e.g., “DNS Zone Transfer is TCP Port 53 not UDP Port 53… These people didn’t even bother to check their sloppy work even after I’d pointed out an obvious error”). In addition, male participants offered more detailed suggestions for course improvement, while female course participants complained more often about technical difficulties (e.g., software installation issues).

Another topic that falls into the theme field Course Content and only appeared in the male reviews is topic *18male*. This topic concerns, on the one hand, the lack of support from the instructor and, on the other hand, the course content not having been updated. The reviews mentioned that the instructor neglected the course and was, therefore, less available for questions and no longer regularly updated the course.

Exam preparation. Topics *12female* and *16male* address the challenge that online courses poorly prepare students for certification exams. The male reviews predominantly concerned poor preparation for cloud computing certifications, while the female reviews were more broadly thematic. In addition, the female reviews contained more criticism regarding spelling, grammar, or typing errors than the male reviews. Topics *13female* and *17male* specifically include criticisms of errors in the courses' practice tests regarding spelling, grammar, and typing mistakes as well as incorrect or misleading test questions or answers.

Finally, we would like to discuss the two remaining topics. Despite careful qualitative analysis of all the most probable words and reviews for both topics, we could not extract any coherent and interpretable themes from topics *14female* and *19male*.

The results of the analysis offer valuable insights into the differing perceptions and experiences of MOOCs among men and women in IT and software courses. Our findings highlight gender-specific expectations and criticisms, indicating important areas for promoting gender inclusivity in online education. The guidelines we derive next address these areas to make MOOCs more gender inclusive. In addition, we discuss the cause of these differences in interviews with experts. This discussion contributes directly to STT by demonstrating practical manifestations of the differences caused by stereotype threat in online IT courses. The results highlight the significance of considering gender as a factor in course design and content delivery, thus addressing RQ1.

5 Expert interviews

This section presents the insights gathered from expert interviews, shedding light on the nuanced factors underlying the gender-based differences observed in our reviews in Sect. 4. We applied the grounded theory approach, which is particularly appropriate for our study because it helps researchers to understand complex social processes (Willig, 2009). Through a precise exploration of expert perspectives, Sect. 5 elucidates the multifaceted dynamics shaping gender-related patterns in online IT and software education. The section also offers valuable contributions to understanding gender inclusivity in online learning environments, paving the way for actionable recommendations to promote equitable educational experiences.

5.1 Methods

A key characteristic of grounded theory research is that, to explore a phenomenon as deeply as possible, data collection and analysis are closely linked. Accordingly, the analysis results of our LDA data influenced the subsequent collection of our interview data, including the selection of interview participants and questions (Glaser et al. 1968). We conducted interviews to complement the results of our LDA analysis of 338,459 MOOC reviews. We conducted 10 semi-structured expert interviews to explain the gender-based differences in the reviews we analyzed. We followed positivist qualitative research and practiced abductive theory building (Dubé & Paré, 2003; Eisenhardt 1989). Although grounded theory is associated with qualitative research, many of its basic assumptions appear

Table 4 Overview of experts interviewed

		Interviewee	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
		Duration [min]	70	60	40	49	59	64	93	40	57	49
		Expert Function	Gender and Language Researcher	Gender Equality Officer	Project Coordinator	Political Scientist	Gender Inequality Researcher	Female Leadership Coach	Professor for Informatics Didactics	Female Leadership Coach	AI Researcher	CEO at a Technology Diversity Association
Expert knowledge		Gender	x	x	x	x	x	x	x	x	x	x
		IT	x		x				x		x	
		Education	x	x	x	x	x	x	x	x		x
Explanations for gender-based differences	IT-specific	Stereotype Threat Theory	x	x	x	x		x	x	x	x	x
		Technical Background					x	x			x	
	IT- independent	Agency and Communion	x	x	x	x		x	x	x	x	x
		Negotiating Demands	x	x				x	x	x		
		Attribution of Mistakes	x							x		x

Table 4 (continued)

	Interviewee	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
	Duration [min]	70	60	40	49	59	64	93	40	57	49
	Expert Function	Gender and Language Researcher	Gender Equality Officer	Project Coordi-nator	Political Scientist	Gender Inequality Researcher	Female Leadership Coach	Profes-sor for Informatics Didactics	Female Leadership Coach	AI Researcher	CEO at a Technology Diversity Association
Designing gender-inclusive online IT courses	Gender-Inclusive Use of Language	x	x	x	x		x	x	x	x	x
	Gender-Inclusive Course Grouping	x		x			x	x	x		x
	Role of the Instructor		x	x	x	x		x		x	
	Opportuni-ties for Commu-nication		x		x	x	x	x		x	x
	Low-Threshold Help Channels				x	x	x	x		x	x
	High Level of Didactic Quality	x	x			x	x	x	x		

Table 4 (continued)

Interviewee	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
Duration [min]	70	60	40	49	59	64	93	40	57	49
Expert Function	Gender and Language Researcher	Gender Equality Officer	Project Coordinator	Political Scientist	Gender Inequality Researcher	Female Leadership Coach	Professor for Informatics Didactics	Female Leadership Coach	AI Researcher	CEO at a Technology Diversity Association
Raising Awareness	x			x						x

in retrospect to be positivist (Charmaz 2014). The grounded theory approach involves collecting data from multiple sources to understand participants' experiences. In our study, we used theoretical sampling for interviews to complement the results of our extensive LDA analysis, including 338,459 MOOC reviews. Theoretical sampling requires that data sources be selected based on their utility in further elucidating the emerging theory, rather than on predetermined quotas for data collection (Murphy et al. 2017). It posits that data collection and analysis should continue until no new information emerges. In our case, interviews informed by LDA results allowed for direct exploration, culminating in theoretical saturation after 10 interviews (Glaser et al. 1968). To fully explore the intersection of education, gender, and IT, we selected interviewees with expertise in at least two of these areas (see Table 4). In addition, we sought diversity in professional background and age to capture evolving perspectives on gender equity across decades. We interviewed individuals from academia and industry to integrate both research and practice perspectives into our findings. A summarization of the metadata concerning the experts' characteristics is provided in Table 5. A key aspect of our theoretical sampling strategy was to interview only women. Given the complex social dynamics underlying the gender disparities observed in the reviews we analyzed, we hypothesized that male interviewees might find it difficult to delve deeply into women's implicit experiences of discrimination in IT education. By focusing on female interviewees, we encouraged experts to continually reflect on their own encounters with such discrimination in IT education.

Interviews lasted 40 to 93 min, were conducted between 05.01.2023 and 27.01.2023, and followed the guidelines of Myers and Newman (2007). To give interviewees a sense of comfort and minimize social dissonance, only the female second author, who has a background in psychology and primarily conducted the interview, and the male first author, who has a background in information systems and assumed the observer role, attended the interviews. Having one male and one female researcher in attendance was intended to avoid gender bias in the analysis results. The female researcher conducted the interviews to provide the female participants with a conversation partner who could easily empathize with the implicit discrimination against women in IT education. With her background in psychology, she was able to sensitively address the issues of discrimination that were raised during the interviews. The aim was to create a safe interview environment and prevent bias in the data, such as socially desirable response behavior. The semi-structured interview guideline derived from the findings of the LDA results allowed us to adapt to the interviewees and aimed to find explanations for the gender-based differences in the reviews we analyzed. In addition, the interview questions were continually adjusted based on the insights gained during the interview process (Glaser et al. 1968). As a result, our in-depth interviews became increasingly focused and tailored as the study progressed. The first questions focused on introducing the interviewees and describing their professional and personal experiences with our research topic. Next, the researcher outlined the LDA results roughly at first and clarified questions to ensure that the researcher and expert were on the same track. Thereafter, the researcher described the LDA results in detail and asked the experts to share their

Table 5 Additional expert metadata

Interviewee	Age	Degree	Years of working experience	Position/Job title	Field
#1	45	Psychology Ph.D	23	Professor	Research
#2	37	Education M.A	11	Deputy Director	Equal Opportunity Office
#3	28	Marketing / Industrial Psychology M.Sc	5	Project Manager	Education-Tech
#4	33	Political Science M.A	8	Researcher	Research
#5	43	Sociology Ph.D	18	Team Leader	Research
#6	52	Business Administration Diploma	28	CEO	Corporate Gender Equality Consulting
#7	50	Computer Science Ph.D	25	Professor	Research
#8	58	Law 2. State Exam	25	CEO	Corporate Gender Equality Consulting
#9	46	Computer Science Ph.D	18	Senior Researcher	Research
#10	61	Sociology Ph.D	22	CEO	Nonprofit for Diversity in Technology

proposed explanations for the gender-based differences in the reviews we analyzed. After each proposed explanation, we asked a follow-up question about what practical guidelines they recommend for the gender-inclusive design of online IT courses, based on their own experiences.

After conducting the interviews, we recorded, transcribed, and abductively coded the interviews. We divided the emerging codes into (1) proposed explanations for the identified gender-based differences in the reviews we analyzed and (2) guidelines on how to design a gender-inclusive online IT course. Following the grounded theory approach, we continuously compared concepts and patterns within our emerging interview data, existing LDA results, and relevant literature to formulate theoretical propositions rooted in both historical and contemporary data, taking into account the existing literature (Strauss & Corbin 1990). In the first round of coding, two researchers assessed the interview data and conducted open coding. They compared individual codes and resolved any discrepancies through discussion until a consensus on appropriate coding was reached. Two additional researchers participated in the second round of coding to increase reliability and further discussed the assignment of codes. The goal was to refine the conceptual framework of our categories and elevate the data to a conceptual level. During this process, the researchers considered the relationship between these categories and existing research. On the basis of this methodological approach, we derived five propositions to explain our identified gender-based differences in the reviews we analyzed (Chapter 5.2.1) and seven guidelines on gender-inclusive online IT course design (Chapter 5.2.2). Table 4 overviews the resulting codes and the number of experts who mentioned them during their interviews.

5.2 Results

5.2.1 Explanations for identified gender-based differences

We divide our hypotheses derived from the interviews into explanations that attribute the cause of the gender-based differences identified in the analyzed reviews to (1) the characteristics of women in IT (IT-specific) and (2) gender-based differences in online learning in general (IT-independent). The expert interviews provide explanations for the identified gender-based differences in MOOC participation and criticism, consistent with STT. They show how stereotypes contribute to the behavior and feedback patterns of men and women in IT courses. The findings thus offer real-life examples of the practical consequences of STT, broadening the understanding of this theory. By addressing both IT-specific and IT-independent factors, we provide a holistic view of the differences between men and women, thus answering RQ1. Additionally, expert recommendations for gender-inclusive MOOC design offer direct guidance on strategies to address these challenges, thus answering RQ2.

IT-specific: stereotype threat theory. STT provides explanations for the gender-based differences in the reviews we analyzed. One example is the dominant writing style we found in the male reviews. For example, male participants were more likely to list only mistakes made by the instructors in their explanations (*11male, 12male, 13male, 14male, 15male*), and male reviews were harsher and more assertive in requesting a refund (*6male*). According to STT, this dominant writing style can be explained by the fact that when men engage with a stereotypically male topic such as IT and software, they are not at risk of fulfilling a negative stereotype and, therefore, feel more comfortable criticizing extensively. By contrast, when women struggle with gender role insecurities, they might adopt a more restrained writing style.

Furthermore, across all topics, male reviews are between 21 and 166% longer than female reviews. Particularly concerning topics with a technical focus (*11female*, *15male*), we noticed that male course participants go into much more detail than female participants do. The average male review is 85% longer than the average female review. The difference is even more pronounced in the topics on explaining coding (*10female, 13male*). Here, female course participants describe more how they would have liked the content to be explained, while male participants list errors. In these topics, the difference in review length is even more apparent, with the average male review being 166% longer than the average female review. Regarding STT, it can be concluded that men feel more confident and are more willing to criticize IT courses than women because they do not risk fulfilling a stereotype or feel better educated in IT. Therefore, the interviewees hypothesized that men will write more confident and, consequently, more detailed reviews when the topic has a more technical focus.

Another identified gender-based difference that can be explained by the STT is the topic *18male* criticizing the lack of support from the instructor, which does not appear in the female reviews. One explanation for this could be the threat that women in IT experience, which causes them to avoid asking for support. Interviewees indicated that women in IT are accustomed to being ignored when they ask for assistance, and they thus rework the course content on their own rather than asking the instructor for additional help. Only if they cannot find the missing information anywhere on their own would they seek help from the instructor. The threat and pressure to avoid embarrassment by asking "stupid" questions may lead women to ask only precise questions. John and Meinel (2020) also found that women are more likely to ask closed and consistent questions in discussion forums and, accordingly, receive less vague answers.

Proposition 1: *As male participants are not at risk of confirming a negative stereotype, they exhibit a more dominant writing style when reviewing online software and IT courses than female participants.*

IT-specific: technical background. The interviewees offered another explanation for the differences we identified: The reviews we compared may differ not only in terms of gender but also in terms of the participants' professional background. The interviewees hypothesized that male course participants are more likely to have technical backgrounds and that female course participants are more likely to enter IT and software courses as career changers. Furthermore, the interviewees hypothesized that female course participants acquire IT competencies as additional qualifications for their current profession, for example as teachers, due to the changes in the wake of digitalization. The tendency for girls not to enter IT-related fields directly after high school is due to stereotypical expectations as well as gender-normative attitudes in students' social media and societal environments. Underlying stereotypes include that boys generally have a more positive self-image in IT-related fields than girls do (Ehrlinger et al. 2018; Sultan et al. 2020). Therefore, such stereotypical thinking may unconsciously influence girls' interests and result in them not entering IT-related fields immediately after graduating from high school but entering IT and software courses later. A more technical professional background among male course participants is one explanation for their tendency toward a more extensive reviewing style. We hypothesize that male participants, due to their technical background, tend to criticize more technical errors (*11male, 12male, 13male, 14male,* and *15male*) and write longer reviews. By contrast, female reviewers criticize technical aspects less and the instructor's didactic teaching style more (*8female, 9female, 10female,* and *11female*).

The more technical background of the male course participants also explains the absence of the support topic in the female reviews, according to the interviewees, who hypothesized that men tend to already have more experience with IT and software MOOCs and are therefore better able to assess the amount of support they can expect from the instructor and are more critical when support is lacking. Women, by contrast, have less experience with IT and software MOOCs, do not know what they are entitled to, and are correspondingly less critical when support is lacking.

Proposition 2: *Compared with female participants, male participants are more likely to have a technical professional background, and they thus display a more dominant writing style when reviewing online software and IT courses.*

IT-independent: agency and communion. Another explanation for the more dominant writing style of the male reviews is what U.S. psychologist Bakan (1966) describes as the "agency" personality trait, which is characterized by attributes such as self-assertion, power-seeking, and achievement and is significantly more prevalent in men. This type of assertiveness is found more in male reviews, as males often only list the mistakes instructors made in their explanations (*11male, 12male, 13male, 14male,* and *15male*). The interviewees hypothesized that by focusing more on the instructor's mistakes, participants with a high agency trait tend to engage in competition with the instructor to improve their position. They also suggested that the longer reviews of men indicate that they want to demonstrate their competence.

By contrast, the personality trait that is significantly more common in women is described by Bakan (1966) as "communion." According to Bakan's definition, communion primarily involves the sense of belonging to a community. According to the interviewees, women's tendency to focus on community explains the gender-based differences in our analyzed reviews. For example, female reviews are notably expressed more politely and less harshly when it comes to asking for a refund (*4female*). In addition, the shorter female reviews indicate a more restrained writing style that leaves room for other members of their community. In addition, female reviews are more constructive and offer more suggestions for improvement. For example, women were more likely to indicate how they would have liked the content explained (*8female, 9female, 10female,* and *11female*), made more suggestions for improvement of the presentation style (*3female*), and provided more specific examples of free alternative courses (*5female*). This constructive criticism and the concrete, specific suggestions for improvement demonstrate that, on average, women are more concerned with the success of the community than with self-promotion.

Proposition 3: *Due to gender socialization, female course participants have a more communal reviewing style, and male course participants have a more agentic reviewing style in online courses.*

IT-independent: negotiating demands. Gender-based differences in the reviews we analyzed indicate that male course participants tend to have higher expectations of courses and express their demands more dominantly than their female counterparts. This is reflected, for example, in their harsh tone when asking for a refund (*6male*) or in the fact that only male reviews reflect a topic that demands further support from the instructor (*18male*). According to the interviewees and previous research, this tendency for men to be more articulate in their demands is rooted in their socialization process in which they learn to negotiate their demands (Dildar & Amjad 2017). In addition, men are more likely to assume financial responsibility due to their socialized role in the family context (Haines et al. 2016). Women, by contrast, tend to learn to withdraw themselves and to be allowed to demand little.

Proposition 4: *Due to gender socialization, female course participants demand less than male course participants when reviewing online courses, as male participants are more likely to learn to negotiate their demands.*

IT-independent: internal and external attribution of mistakes. Gender-based differences in the reviews we analyzed can also be explained by the fact that these differences exist in the attribution of mistakes. While men tend to attribute mistakes more externally, women tend to attribute mistakes to internal causes due to more pronounced self-doubt (Koch et al. 2008). The interviewees suggested that he tendency of men to attribute mistakes to external causes could explain why they attribute mistakes to the instructors and therefore more frequently only list errors made by the instructors in their explanations (*11male, 12male, 13male, 14male,* and *15male*). The female reviews, by contrast, criticize the didactic style of explanations more constructively (*8female, 9female, 10female,* and *11female*) and offer direct suggestions for improving the instructors' presentation skills (*3female*).

Proposition 5: *Due to gender socialization, female course participants attribute mistakes to internal causes when reviewing online courses, whereas male course participants attribute mistakes to external causes.*

For a clearer picture, we have initially presented the hypotheses separately, as they are refutable hypotheses that can be tested both in isolation and in combination using deductive research methods. In reality, however, we assume interdependencies between the hypotheses. Mapping the propositions to the areas of gender-based differences demonstrates the interconnected nature of the propositions, underscoring their collective role in illustrating the complexity of gender dynamics in online computer science courses (Table 6). We adopted this mapping approach to highlight that these propositions, while offering distinct insights, are not entirely separate entities but, rather, jointly influence students' behavior and experiences. In line with the experts, our propositions are based on the assumption that gender-based differences are not caused by a single, isolated factor but are the result of a complex interplay

Table 6 Mapping of propositions and areas of gender-based differences

Proposition Area of gender-based difference	Stereotype threat theory	Technical background	Agency and communion	Negotiating demands	Internal and external attribution of mistakes
Communication style and feedback culture	X		X		X
Asking for support	X	X		X	
Exam preparation			X		

of factors. Acknowledging the interconnectedness of these insights allows us to present a more nuanced and holistic view of the gender dynamics within IT and online learning environments. The overlapping theoretical themes are in line with the grounded theory approach and enrich the analysis by capturing the complexity of the phenomena studied (Glaser et al. 1968). This perspective reinforces the importance of considering the propositions in this study not as isolated variables but as parts of a larger, interconnected framework that collectively offers a heightened understanding of the gender gap in computer science education.

The relevance of each proposition to a specific area of gender-based difference is rooted in the interplay of socialization, stereotypes, and individual backgrounds. In the area of communication style and feedback culture, stereotype threat theory (proposition 1) explains why male participants may adopt a more dominant critique style, feeling less constrained by the fear of fulfilling negative stereotypes. Female participants, by contrast, may internalize mistakes and offer more constructive feedback, aligning with the fifth proposition on the internal and external attribution of mistakes. The difference in agency and communion (proposition 3) further delineates these communication styles, with men's agency leading to assertiveness and women's communion leading to a focus on community and constructive suggestions. With regard to asking for support, STT (proposition 1) suggests that women might hesitate to seek support, fearing stereotype confirmation, whereas men, having a technical background (proposition 2), confidently express their need for support, as they understand the specific assistance they require. Men's socialization to negotiating demands (proposition 4) also makes them more vocal in critiquing support, a behavior less observed in women due to a different form of socialization. Exam preparation is particularly linked to the agency and communion proposition (proposition 3), where the communal aspect of learning is emphasized. Women's preference for detailed explanations suggests a more communal approach to learning, wherein understanding and collaboration are valued, which contrasts with the more solitary or competitive preparation strategies of men.

5.2.2 Designing gender-inclusive online IT courses

In addition to our theoretical contribution, we also derived guidelines for practitioners from the interviews. Anyone involved in the instructional design process can implement them to design gender-inclusive online IT courses.

Gender-inclusive and practice-oriented use of language. Interviewees emphasized the importance of using gender-inclusive language in online IT and software courses when both male and female participants engage. This includes the balanced representation of male and female characters in texts and images to allow learners to identify with characters of their gender and thus create role models. Another way to use gender-inclusive language and incorporate role models is to cite both male and female authors as much as is possible. In addition, interviewees noted the implicit effect of gender stereotypes in textual and visual learning content on learners' self-efficacy. By gender-stereotypical language use, we mean, for instance, using stereotypical male examples such as dragons and castles to vividly explain learning content. Prior research has addressed this use of language by studying, for example, communal and agentic words in job advertisements (Gaucher et al. 2011) or gender-stereotypical language in textbooks (Göritz et al. 2022). According to the interviewees, reducing the atmosphere of masculinity in IT and software courses by using gender-inclusive language could not only encourage more women to engage with IT and software learning content but also cause men to feel less pressured to behave in an agentic manner by consistently demonstrating their competence.

Interviewees also indicated that when using examples in IT-related courses, it is important not only to pay attention to gender-inclusive language but also to use practice-oriented examples to illustrate the benefits of IT and software. By emphasizing the practical benefit of IT and software, learning content in this domain is perceived as less stereotypically IT and, consequently, as less stereotypically male. However, gender-inclusive and practice-oriented language is essential not only within the learning content but also before and after the course, for example in course marketing or within support channels.

Gender-inclusive course grouping. Previous research has explored whether male and female participants in IT- and software-related courses should be grouped (Busch 1996; Paloheimo & Stenman 2006). On the one hand, some interviewees argued that having a course with only female participants can foster a trusting atmosphere where women can address their self-doubts and feel less alone. Moreover, in courses with exclusively male or female participants, fewer gender-stereotypical roles may emerge during group work, such as the men programming while the women take notes. On the other hand, other interviewees argued that mixed-gender IT and software courses prepare female participants for the real world, in which women must learn to assert themselves. When teaching mixed-gender groups, attention should be paid to the emergence of gender-stereotypical roles, such as during group work. A course system could, for example, automatically assign course participants rotating roles and responsibilities, thereby allowing female course participants to assume unfamiliar roles, which would increase the chance of positive experiences that could lead to enhanced self-efficacy. In addition, interviewees noted that in IT courses where women are underrepresented, instructors should group all female participants. For example, instead of assembling three groups with 15% women, they should aim for two courses with only men and one course with 50% women. Reducing women's feelings of loneliness in a counter-stereotypical IT and software course can reduce the triggering of stereotype threat.

Role of the instructor. Interviewees also considered the allocation of instructors important for a gender-inclusive learning environment. Promoting instructors with diverse characteristics on an online learning platform makes it possible to target different types of learners. Here, the instructor's gender also plays an important role, as the instructor often acts as a role model and provides room for identification while the participant engages with IT and software learning content. The interviewees emphasized that this is not limited to female instructors teaching female participants, as male instructors who teach in a didactically sound and gender-inclusive way can also serve as role models for all students.

Opportunities for communication. Interviewees emphasized that in digital learning environments, providing opportunities for participants to communicate and interact with one another is essential. In this context, it is also important to avoid creating an atmosphere of masculinity in communication channels, as it would trigger stereotype threat. Interviewees particularly called for gender-inclusive discussion forums on online learning platforms that, for example, promote the use of gender-inclusive language. They also emphasized the need for well-moderated forums that prevent, for example, hate speech, wiseacres, and mansplaining. One way to support moderation in discussion forums through technology would be to automatically detect when a participant's questions are systematically ignored by other participants or by the instructor. To avoid stereotype threat, discussion forums could also be pseudonymized, such that participants and instructors cannot presume the gender of other participants based on their usernames. In this context, it is important to note that while pseudonymization serves to avoid stereotype threat, users can still be traced in case of misbehavior. Interviewees also highlighted that a balance between pseudonymization and positive experiences of personal interaction should be a key target in the design of gender-inclusive discussion forums.

Low-threshold help channels. The interviewees hypothesized that female participants experience uncertainty due to a lack of self-efficacy in the IT context and, therefore, rework course content on their own rather than asking the instructor for additional help. To reduce this psychological barrier, the interviewees proposed offering low-threshold help channels and explicitly addressing them as part of the courses. Help can be provided, for example, through anonymous Q&As, chatbots, or fixed time slots for a live chat with the instructor. In direct conversations with the instructor, creating an atmosphere that reduces uncertainty when asking for help is important. Concerning chatbots, they offer the advantage that participants do not have to ask a real person for help, which might reduce insecurities and fear of prejudice. In addition, chatbots can use NLP techniques to automatically adapt their communication style to the participant asking the question, for instance by detecting uncertainties in the participant's writing style and consequently providing more detailed and easier-to-understand answers. The interviewees emphasized that chatbots must operate in a context-sensitive manner. If a chatbot gives nonsensical answers, this can result in a high level of frustration among course participants and a high dropout rate, especially among those who are already uncertain.

High level of didactic quality. To avoid stereotype threat in IT and software courses, interviewees further recommended attentive implementation of didactic principles, which includes the division of the course content into small learning units, the self-directed acquisition of learning content by participants, and individual adaptation of the learning pace. Attentive implementation of didactic principles leaves little room for uncertainty among course participants and, consequently, for gender bias. Interviewees particularly emphasized the importance of the learning principal of individualization. They stated that it is necessary to ensure that every course participant's learning channel, regardless of gender, is addressed. Furthermore, the level of difficulty should be adjusted within each course to avoid overwhelming participants who already struggle with insecurities. For further individualization, instructors can, for example, enquire about the course participants' prior knowledge, demands, and expectations. In this context, it is important to keep the format anonymous to obtain unbiased feedback.

Raise awareness. The last aspect mentioned by interviewees for designing gender-inclusive online IT courses is raising awareness among everyone involved in the instructional design process. Already during the design of the online course, for example, platform providers or course instructors run the risk of unconsciously incorporating gender stereotypes into the course, which in turn can trigger stereotype threat (Stumpf et al. 2020). Interviewees emphasized the need to provide the individuals involved in designing online IT courses with opportunities to learn about unconscious bias to raise awareness of gender-specific challenges in online IT courses. Not only for the platform providers and course instructors but also for the course participants, directly addressing discrimination and stereotypes at the right time can raise awareness and overcome stereotype threat.

6 Discussion

Based on our research questions, we investigated the different challenges that men and women face when participating in online IT courses. We conducted expert interviews to find explanations for these differences and divided our propositions about gender-based differences in MOOC reviews into two categories. The first, "IT-specific," highlights differences arising from women's involvement in the traditionally male-dominated field of IT and software. The second, "IT-independent," examines differences arising from gendered behaviors independent of learning content. This comprehensive approach, which includes both IT-specific and IT-independent factors, addresses RQ1 by exploring, through data and expert interviews, how and why men and women face different challenges in online IT courses. We further developed guidelines for designing online IT courses that enable every user, regardless of their characteristics, to participate with confidence. By addressing the causes of differences in online learning environments, these guidelines answer RQ2 by demonstrating how the gender-inclusive design of MOOCs can contribute to overcoming the gender gap in computer science.

Our research contributes to the discourse on STT by extending its application to the domain of online education, specifically MOOCs. Our analysis of negative user reviews from Udemy revealed the different issues that men and women face. In discussions with experts, we identified stereotype threat as a potential reason for these differences. Therefore, we present examples of how gender stereotypes affect learners' experiences in the real world. Specifically, we highlight the engagement and performance issues that female participants face. These results connect the theoretical foundation of STT to data-based outcomes in digital education environments, showing how stereotype threat is relevant in the context of online learning. Based on these identified differences, we propose guidelines for the gender-inclusive design of MOOCs. These guidelines aim to counteract the challenges identified, enhancing inclusivity and accessibility in online IT courses. Our work serves as a practical blueprint in the existing theoretical discourse on STT, proposing effective strategies to bridge the gender gap in computer science education and beyond.

For the information systems (IS) discipline, we extend the domain-specific knowledge of information systems design, particularly the design of platforms, by applying a theory from the social sciences. Thereby, we demonstrate the need for interdisciplinary research on original IS topics such as platform design by incorporating insights and knowledge from other disciplines, such as psychology, gender studies, and education. Previous IS research has rarely used STT as a kernel theory for inclusive design, and in the past, it has been addressed only in a small number of IS journals (e.g., Computers in Human Behavior). However, the topic of inclusion is currently rapidly gaining prominence in the IS community, with tracks emerging at IS conferences, such as AMCIS23 and ECIS23, calling for submissions on the topic of social inclusion and the social and ethical implications of using information and communications technology (ICT). We recognize the need for more inclusive information systems and, therefore, use STT to infer drivers of gender inequalities in ICT use in learning contexts to lay the groundwork for inclusive digital learning environments. Furthermore, our research provides a practical contribution through clear guidelines that practitioners can apply to make their courses more gender inclusive. Finally, our guidelines expand previous research on the inclusive design of digital learning environments with further insights.

One such guideline the *Gender-Inclusive and Practice-Oriented Use of Language* relates to the importance of language when male and female students participate together in IT and software courses. Gender-inclusive language is often associated with masculine generic constructions by using masculine nouns and pronouns when referring to people in general. However, our results highlight the occurrence of implicit discrimination in the form of stereotypical language in learning content. Cheryan et al. (2011) investigated the elimination of gender stereotypes in learning environments, demonstrating that a gender-neutral learning environment increases women's interest in computer science. Another approach to addressing implicit discrimination is to present diverse verbal and visual cues in the form of examples, shapes, and colors, such that all students, regardless of gender identity, can identify with the learning content (AlSulaiman & Horn 2015). The attentive use of verbal and visual cues in learning environments is important for IS research, as sociotechnical systems such as MOOCs reach millions of learners with just a few clicks. Thus, further IS research on the design of inclusive digital learning environments would make a significant contribution not only to theory but also to society in general. An automated approach to implementing inclusive language in learning contexts was previously discussed in the IS discipline, for example by Göritz et al. (2022), who automatically identified gender stereotypes in learning content.

Using gender-inclusive language is important not only for learning content but also at other points in the learning process, such as for communication channels. As stated in our guideline *Opportunities for Communication*, communication channels play a crucial role in ensuring gender-inclusive MOOCs. Previous research has shown that communication channels such as moderated online discussion forums strengthen students' sense of community and, consequently, their resilience (Richard & Hoadley 2015). Pseudonymizing these forums can further reduce concerns about social perceptions and, hence, lower stereotype threat (Lee et al. 2017). Furthermore, as described in our guideline *Low-Threshold Help Channels,* previous research has found that agent-based learning technologies can provide low-threshold help for uncertain students who are afraid to ask for it due to a lack of self-efficacy in the IT context (Leider & Strobel 2020; Rosenberg-Kima et al., 2010). Within those channels of communication, the use of inclusive language is important to prevent stereotype threat. IS research now faces the need to investigate new approaches to providing inclusive digital communication channels on learning platforms. We call on IS researchers to examine how, for instance, forums and chatbots on learning platforms can be designed to be inclusive by using STT. One example of how to implement technology-driven moderation in discussion forums would be to automatically detect when a participant's questions are systematically ignored by other participants or the instructor.

Another guideline for practitioners to design a learning environment that promotes confidence-building learning experiences is the *High Level of Didactic Quality*, which suggests the careful implementation of didactic principles, especially the principle of individualization. Individualization here refers to the customization of MOOC design elements to learners based on their belonging to an underrepresented group (Harris et al. 2020). In this context, cross-disciplinary discussions address, for example, whether teachers should be assigned based on their characteristics in relation to students' characteristics (e.g., gender). As discussed in our guideline *Role of the Instructor,* interviewees emphasized that including instructors with different characteristics on an online learning platform makes it possible to target different types of learners. Kizilcec et al. (2020) further note that greater diversity among instructors provides the opportunity to present a variety of role models. However, this does not apply solely to female instructors teaching female participants, as male instructors who teach in a didactically sound and gender-inclusive way can also serve as role models for all students.

A similar question arises when considering whether gender-specific or gender-mixed course groups are preferable. Previous research has shown that the presence of more women in courses in male-dominated fields such as IT and software leads to significantly higher engagement and discussion behaviors (Brooks et al. 2018) and fewer experiences of sexism (Ypma 2019) among women. However, as stated in our guideline *Gender-Inclusive Course Grouping*, some interviewees argued for gender-mixed IT and software courses because they prepare participants for the real world, where women must learn to assert themselves. The answer to the question of whether gender-specific or gender-mixed course groups are preferable leads to different outcomes for different disciplines. Thus, to answer this question holistically, it is necessary to further examine it through the lens of IS research.

Overall, our guidelines contribute to the IS body of knowledge on the inclusive design of information systems, especially of digital learning platforms. Designing information systems that prevent stereotype threat is essential in providing an attentive user experience for those who would otherwise experience uncertainty because they belong to an underrepresented group. Consistent with previous research and summarized in the guideline *Raise Awareness*, interviewees emphasized the importance of raising awareness among all individuals involved in the instructional design process (Harris et al. 2020; Ypma 2019) to ensure that gender stereotypes are not unconsciously implemented into course design (Stumpf et al. 2020).

We provide online course designers with specific guidelines for making their courses gender inclusive. Implementing our recommended guidelines into their learning platforms would increase the quality of their courses for previously underrepresented user groups while maintaining the same quality for existing user groups. This would lead to an increase in the total number of course participants and, consequently, to higher revenues and profits for course designers since online courses are easily scalable. The inclusive design of digital learning environments can also be used to promote the courses and retain existing users.

In addition to the economic benefits for individual organizations, our research findings are also of societal importance, as they contribute to the European Union's sustainability goals and the European Commission's education goals. The EU defines "gender equality" under point 5 of its sustainability goals, with the intention to end all forms of discrimination, violence, and harmful practices against women and girls. The EU mentions quality education as a key objective to achieve this equality.[8] The European Council (2021) further defines that "inclusive education and training also entails developing gender sensitivity in learning processes and in education and training institutions, and challenging and dissolving gender stereotypes, especially those that constrain the choices of boys and girls for their field of study."

The practical guidelines we recommend contribute to a high-quality learning experience. Designing gender-inclusive digital learning environments in stereotypical subject areas such as STEM or nursing promotes diversity in the labor market. As a result, companies will have a larger and more diverse pool of applications from which to choose in the future, which in turn will help address societal challenges such as the gender pay gap or the labor shortage. From this perspective, the gender pay gap is a key example in our specific use case of computer science education. One of the main reasons for the gender pay gap is the difference in pay between male-dominated and female-dominated fields: On average, female-dominated fields pay less than male-dominated fields (Wrohlich 2017). Through the implementation of our guidelines, more women could be motivated to enter and succeed in computer science, which has traditionally been a male-dominated and high-paying sector of the economy. They could thus improve their earnings, which would narrow the gender pay gap. This effect would be further strengthened by increasing the number of women in computer science, which in turn would have a pull effect, attracting more women into computer science (Chhaochharia et al. 2022; Friedrich et al. 2019). A similar effect can be applied to female-dominated professions such as nursing, which are still poorly paid on average and are affected by a shortage of skilled workers (Caputo & Ross 2022; Yip et al. 2021). Inclusive course design could help men to identify more strongly with the content of nursing care and thus work in this field.

Overall, we recommend applying our guidelines to any field that presents a gender-biased picture of its workforce. These fields may be affected by stereotype threat and insecurities among potential employees, preventing them from reaching their full potential. For example, girls and women may face societal expectations based on their gender as they pursue an education to become programmers. This is a problem not only for the individual but also for society, as the IT industry is already

[8] https://ec.europa.eu/eurostat/web/sdi/database/gender-equality, last accessed on 06.07.2023.

suffering from a labor shortage. This issue can only be addressed through higher quality and more inclusive education, as digitalization will continue to change the labor market's demand for more highly skilled IT workers.

7 Conclusion

Our research demonstrates that gender-based differences exist in learning in MOOCs. Some of these differences are domain-independent, while others are specific to computer science courses. We emphasized the importance of recognizing these differences, investigated how they arise, and provided guidelines for addressing them through the design of gender-inclusive online learning environments. These differences are influenced by factors such as stereotype threat, level of experience with technology, gender socialization, negotiation strategies, and attribution of mistakes. Stereotype threat theory played a central role in our study. Our findings suggest that stereotype threat affects the performance and engagement of female learners in computer science courses. This highlights the need for MOOC providers to pay attention to gender-inclusive course design that does not trigger stereotype threat.

Our research extends IS knowledge by applying a social science theory, namely stereotype threat theory, to the design of information systems, specifically online education platforms, highlighting the value of interdisciplinary approaches. Our findings have far-reaching implications. For MOOC providers, course designers, and instructors, our research provides a roadmap for designing gender-inclusive learning environments; for scholars, it provides hypotheses and explanations for the influence of gender on the design of information system artifacts; and for policymakers, it underscores the need for policies that promote diversity and inclusion in online education.

Due to the large amount of user data, we expect a high generalizability of our results. However, there are some limitations to our approach that should be considered in future research. First, we could only analyze a subset of the top 10,000 courses and ratings on Udemy due to technical limitations in accessing ratings and courses. The limitation on the number of ratings extracted only affected the top 47 courses and is therefore negligible and possibly even beneficial, as these courses with more than 10,000 ratings are not overrepresented in the analysis. Second, we only considered reviews written in English. To increase the generalizability of our results, it would be advisable to include more reviews in the future, especially from other MOOC providers. Nevertheless, given the diversity of the courses analyzed and Udemy's market position as one of the largest MOOC providers, we expect our results to be highly transferable. Third, we had to predict the gender of users based on abbreviations and names. To minimize potential errors when predicting gender based on abbreviations, gender-neutral titles (e.g., Dr.) were excluded from this classification process. While this method may still introduce some inaccuracies, we assume these to be minimal and unlikely to significantly affect our results. To ensure this, we checked random samples of the predicted genders both manually and against the names, following the methodology used for the names. For the names, unidentifiable and unisex categories were included to minimize error, but misclassification could not be completely avoided. However, since the tool we employed had a precision rate of 93% in previous studies (Vasilescu et al. 2014), we believe that on average sufficient precision was achieved. Nonetheless, to increase the potential for gender equality from large, anonymized datasets, future research should explore additional gender identification methods that address both the challenges of gender metadata limitations and privacy concerns and expand the scope and data sources for gender analysis in research. Fourth, our approach only allowed us to make assumptions about binary gender. Therefore, our results do not leave room for other identifications, such as non-binary or genderqueer. It would have been easier to interpret the data and derive hypotheses if we had more information about the users who wrote the MOOC reviews. Fifth, incorporating gender information prior to topic detection may have introduced some bias in the analysis, as the awareness of gender could potentially influence the identification and interpretation of topics. However, this approach was chosen based on grounded theory to ensure conceptual proximity to the social processes under investigation. The inclusion of gender data allowed for a more nuanced exploration of gender-related dynamics within the reviews, which was essential for a deep examination of the phenomenon. Sixth, while the initial separation of data into male and female clusters was strategically chosen to mitigate the effects of a skewed gender distribution on topic detection, it may introduce biases in topical representation. The decision for separation was informed by a study by Puyol-Antón et al. (2021), which demonstrated that segmenting data by demographic groups could enhance fairness and accuracy in the analyses of imbalanced datasets. Despite this strategic approach, the separation by gender may affect the comparability of results across genders, as a greater number of topics were identified in male reviews than in female reviews. However, since topics identified in female reviews were often aggregations of several distinct male topics (e.g., video and audio quality), they still provide a robust foundation for cross-gender content comparison. To further refine our understanding and reduce potential biases, future studies should explore a broader range of analysis methodologies. This includes considering various clustering algorithms and adopting alternative analytical approaches that do not initially segment data by gender, thereby allowing a robust comparison with our results. Finally, the interpretation of LDA is subjective due to the qualitative nature of the analysis. To minimize this bias and increase construct validity through multiple data sources, 10 experts discussed and interpreted the results.

In this study, we examined data from IT and software courses. Our resulting guidelines are specific to women in IT education. We believe that our findings are applicable to other domains where different underrepresented groups are prevalent (e.g., men in social professions), but further research is required to validate this transferability. Future research should also increase the generalizability of our findings by examining user data from other learning platforms or course subjects. Moreover, our guidelines should be implemented and evaluated in additional studies, with a particular focus on an overarching solution that implements all guidelines. Previous work has primarily explored a single-solution approach. A holistic solution may provide more insight into which of the individual guidelines will most benefit an underrepresented group of learners without negatively impacting the overrepresented group. Overall, our findings confirm that e-learning formats such as MOOCs can impart computer science skills in a gender-inclusive way if they are designed accordingly.

In conclusion, our research represents an important step in ongoing efforts to overcome the gender gap in online education and, in the long term, in the labor market. By providing a comprehensive set of guidelines and shedding light on the underlying factors that influence gender-based differences, we are following the trend of the feminist HCI community, transferring it to IS research, and contributing to the design of more inclusive and empowering learning environments for all students.

Acknowledgements The authors would like to thank Dana Dix and Berna Eybey for their help in interpreting the results of the LDA analysis, and all participants of the interviews for their valuable contributions, as well as the reviewers for their constructive feedback.

Author contributions Daniel Stattkus conceptualized the paper. He retrieved the reviews from the MOOC provide Udemy and carried out the data preparation process. He was involved in the interpretation of the results of the LDA analysis and conducted the expert interviews. He was the main author of the data preparation part of the Methodology chapter, the discussion, and the conclusion and co-authored the results. All other chapters were corrected, revised, and aligned by Lorena Göritz conducted the LDA analysis and was involved in the interpretation of the results. She took the leading role in conducting and analyzing the expert interviews. She was primarily responsible for the chapters on Stereotype Threat Theory, Gender-Inclusive Design of MOOCs, and the Results and co-authored the Introduction, Discussion, and the Conclusion. All other chapters were corrected, revised, and aligned by her. Katharina-Maria Illgen was involved in the interpretation of the results of the LDA analysis. She was the main author of the Introduction. She contributed to the theory chapters on Stereotype Threat Theory and Gender Inclusive Design of MOOCs. All other chapters were corrected, revised, and aligned by her. Dr. Jan Heinrich Beinke assisted in placing the paper in the context of previous IS research, provided feedback and suggested revisions throughout the research and writing process. He helped to conceptualize and reflect on the methodological approach. All chapters were corrected, revised, and aligned by him. Prof Dr. Oliver Thomas assisted in placing the paper in the context of previous IS research, provided suggestions for the revision of the article. He helped to conceptualize and reflect on the methodological approach. All chapters were corrected, revised, and aligned by him.

Funding The authors received no funding for the research, authorship and publication of this article.

Declarations

Conflict of interest The authors have no conflicts of interest to declare.

References

AlSulaiman S, Horn MS (2015) Peter the fashionista? computer programming games and gender oriented cultural forms. In: Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play, 185–195.

Amendola M, Castillo C, Passarella A, Perego R (2024) Understanding and addressing gender bias in expert finding task. ArXiv Preprint ArXiv:2407.05335.

Appel M, Kronberger N, Aronson J (2011) Stereotype threat impairs ability building: effects on test preparation among women in science and technology. Eur J Soc Psychol 41(7):904–913

Arun R, Suresh V, Madhavan CEV, Murthy MNN (2010) On finding the natural number of topics with latent dirichlet allocation: Some observations. Pacific-Asia Conference on Knowledge Discovery and Data Mining, 391–402.

Ashcraft C, McLain B, Eger E (2016) Women in tech : The facts. National Center for Women Technology (NCWIT), Colorado

Astleitner H, Steinberg R (2005) Are there gender differences in web-based learning? An integrated model and related effect sizes. AACE Rev (Former AACE J) 13(1):47–63

Bakan D (1966) The duality of human existence: An essay on psychology and religion. Rand Mcnally

Bardzell S, Bardzell J (2011) Towards a feminist HCI methodology: social science, feminism, and HCI. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 675–684.

Baumer EPS, Mimno D, Guha S, Quan E, Gay GK (2017) Comparing grounded theory and topic modeling: extreme divergence or unlikely convergence? J Am Soc Inf Sci 68(6):1397–1410

Bhargava A (2002) Gender bias in computer software programs: a checklist for teachers. Inf Technol Child Educ Annu 2002(1):205–218

Blakemore JEO, Centers RE (2005) Characteristics of boys' and girls' toys. Sex Roles 53(9):619–633

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. J Mach Learn Res 3:993–1022

Brooks C, Gardner J, Chen K (2018) How gender cues in educational video impact participation and retention. International Society of the Learning Sciences, Inc.[ISLS].

Brooks C, Kizilcec RF, Dowell N (2020) Designing Inclusive Learning Environments. In: Proceedings of the Seventh ACM Conference on Learning @ Scale, 225–228. https://doi.org/10.1145/3386527.3405935

Burnett M, Stumpf S, Macbeth J, Makri S, Beckwith L, Kwan I, Peters A, Jernigan W (2016) GenderMag: a method for evaluating software's gender inclusiveness. Interact Comput 28(6):760–787

Busch T (1996) Gender, group composition, cooperation, and self-efficacy in computer studies. J Educ Comput Res 15(2):125–135

Cain CC, Trauth EM (2013) Stereotype threat: the case of black males in the IT profession. In: Proceedings of the 2013 Annual Conference on Computers and People Research, 57–62.

Cao J, Xia T, Li J, Zhang Y, Tang S (2009) A density-based method for adaptive LDA model selection. Neurocomputing 72(7–9):1775–1781

Caputo T, Ross JG (2022) Male nursing students' experiences during prelicensure education: an integrative review. Nurs Educ Today 121:105671

Charmaz K (2014) Constructing Grounded Theory. Sage Publications

Cheryan S, Meltzoff AN, Kim S (2011) Classrooms matter: the design of virtual classrooms influences gender disparities in computer science classes. Comput Educ 57(2):1825–1835. https://doi.org/10.1016/j.compedu.2011.02.004

Cheryan S, Plaut VC, Handron C, Hudson L (2013) The stereotypical computer scientist: gendered media representations as a barrier to inclusion for women. Sex Roles 69(1):58–71

Chhaochharia V, Du M, Niessen-Ruenzi A (2022) Counter-stereotypical female role models and women's occupational choices. J Econ Behav Organ 196:501–523

Clarke N, Mondisa J-L, Packard BW-L, Queener Schemanske C, Tuladhar A, Gosha K (2023) Examining the role of computing identity in the computing experiences of women and racially minoritized undergraduates: a literature review. J Comput High Educ 36:773–812

Clayton KL, Von Hellens LA, Nielsen SH (2009) Gender stereotypes prevail in ICT: a research review. In: Proceedings of the Special Interest Group on Management Information System's 47th Annual Conference on Computer Personnel Research, 153–158.

Cooper J (2006) The digital divide: the special case of gender. J Comput Assist Learn 22(5):320–334

European Council (2021) Council Resolution on a strategic framework for European cooperation in education and training towards the European Education Area and beyond (2021–2030).

Crues RW, Henricks GM, Perry M, Bhat S, Anderson CJ, Shaik N, Angrave L (2018) How do gender, learning goals, and forum participation predict persistence in a computer science MOOC? ACM Trans Comput Educ. https://doi.org/10.1145/3152892

Datascience@berkeley (2021). Changing the Curve: Women in Computing. https://ischoolonline.berkeley.edu/blog/women-computing-computer-science/#:~:text=The proportion of women receiving,bachelor's degree recipients were women

Deveaud R, Sanjuan E, Bellot P (2014) Accurate and effective latent concept modeling for ad hoc information retrieval. Doc Numér 17(1):61–84

Devine PG (1989) Stereotypes and prejudice: their automatic and controlled components. J Pers Soc Psychol 56(1):5

Dildar S, Amjad N (2017) Gender differences in conflict resolution styles (CRS) in different roles: a systematic review. Pak J Soc Clin Psychol 15(2):37–41

Dubé L, Paré G (2003) Rigor in information systems positivist case research: current practices, trends, and recommendations. MIS Q 27:597–636

Dubois PMJ, Maftouni M, Chilana PK, McGrenere J, Bunt A (2020) Gender differences in graphic design q&as: how community and site characteristics contribute to gender gaps in answering questions. In: Proceedings of the ACM on Human-Computer Interaction, 4(CSCW2), 1–26.

Easterbrook MJ, Hadden IR (2021) Tackling educational inequalities with social psychology: identities, contexts, and interventions. Soc Issues Policy Rev 15(1):180–236

Easterbrook MJ, Hadden IR, Nieuwenhuis M (2019) Identities in context: how social class shapes inequalities in education. In: Jetten J, Peters K (eds) The Social Psychology of Inequality. Springer, Cham

Ehrlinger J, Plant EA, Hartwig MK, Vossen JJ, Columb CJ, Brewer LE (2018) Do gender differences in perceived prototypical computer scientists and engineers contribute to gender gaps in computer science and engineering? Sex Roles 78:40–51

Eisenhardt KM (1989) Building theories from case study research. Acad Manag Rev 14(4):532–550

Fedorowicz J, Vilvovsky SG, Golibersuch AJ (2010) Gender differences in teenagers' elective use of computer technology. Commun Assoc Inf Syst 27(1):3

Fiesler C, Morrison S, Bruckman AS (2016) An archive of their own: a case study of feminist HCI and values in design. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 2574–2585.

Friedrich J-D, Hachmeister C-D, Nickel S, Peksen S, Roessler I, Ulrich S (2019) Frauen in IT: Handlungsempfehlungen zur Gewinnung von Frauen für Informatik.

Gaucher D, Friesen J, Kay AC (2011) Evidence that gendered wording in job advertisements exists and sustains gender inequality. J Pers Soc Psychol 101(1):109

Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell 6:721–741

Glaser BG, Strauss AL, Strutzel E (1968) The discovery of grounded theory; strategies for qualitative research. Nurs Res 17(4):364

Göritz L, Stattkus D, Beinke JH, Thomas O (2022) To Reduce Bias, You Must Identify It First! Towards Automated Gender Bias Detection.

Griffiths TL, Steyvers M (2004) Finding scientific topics. Proc Natl Acad Sci 101(suppl 1):5228–5235

Guo Y, Barnes SJ, Jia Q (2017) Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent Dirichlet allocation. Tour Manag 59:467–483

Haines EL, Deaux K, Lofaro N (2016) The times they are a-changing… or are they not? A comparison of gender stereotypes, 1983–2014. Psychol Women Q 40(3):353–363

Harris BN, McCarthy PC, Wright AM, Schutz H, Boersma KS, Shepherd SL, Manning LA, Malisch JL, Ellington RM (2020) From panic to pedagogy: using online active learning to promote inclusive instruction in ecology and evolutionary biology courses and beyond. Ecol Evol 10(22):12581–12612

Hilton JL, Von Hippel W (1996) Stereotypes. Annu Rev Psychol 47(1):237–271

Hines M, Pasterski V, Spencer D, Neufeld S, Patalay P, Hindmarsh PC, Acerini CL (2016) Prenatal androgen exposure alters girls' responses to information indicating gender-appropriate behaviour. Philos Trans R Soc B: Biol Sci 371(1688):20150125. https://doi.org/10.1098/rstb.2015.0125

Hutto C, Gilbert E (2014) Vader: a parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 8, No. 1, pp. 216–225).

Jarvis SN, Nguyen CQ, Zhu M, Ebersole CR, Kray LJ (2023) Do virtual environments close the gender gap in participation in question-and-answer sessions at academic conferences? In search of moderation by conference format. Sex Roles 89(11):818–833

John C, Meinel C (2020) Learning behavior of men and women in MOOC discussion forums–a case study. In: 2020 IEEE Global Engineering Education Conference (EDUCON), 300–307

Kansal AK, Gautam J, Chintalapudi N, Jain S, Battineni G (2021) Google trend analysis and paradigm shift of online education platforms during the COVID-19 pandemic. Infect Dis Rep 13(2):418–428. https://doi.org/10.3390/idr13020040

Kelleher C, Pausch R, Kiesler S (2007) Storytelling alice motivates middle school girls to learn computer programming.In: Conference on Human Factors in Computing Systems - Proceedings, 1455–1464. https://doi.org/10.1145/1240624.1240844

Kizilcec RF, Saltarelli AJ, Reich J, Cohen GL (2017b) Closing global achievement gaps in MOOCs. Science 355(6322):251–252

Kizilcec RF, Saltarelli AJ (2019a) Can a diversity statement increase diversity in MOOCs? Proceedings of the Sixth (2019) ACM Conference on Learning@ Scale, 1–8.

Kizilcec RF, Saltarelli AJ (2019b) Psychologically inclusive design: cues impact women's participation in STEM education. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1–10). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300704

Kizilcec RF, Davis GM, Cohen GL (2017) Towards equal opportunities in MOOCs: affirmation reduces gender & social-class achievement gaps in China. In: Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale, 121–130.

Kizilcec RF, Saltarelli AJ, Bonfert-Taylor P, Goudzwaard M, Hamonic E, Sharrock R (2020) Welcome to the course: early social cues influence women's persistence in computer science. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–13.

Koch SC, Müller SM, Sieverding M (2008) Women and computers. Effects of stereotype threat on attribution of failure. Comput Educ 51(4):1795–1803

Lee M, Starr-Mitchell K, Nunes L, Black M, Schmidt T (2017) MOOCs as facilitator: online learning and women in STEM. In: 2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC), 482–486.

Leider A, Strobel A (2020) Using self-confidence and identity to build perseverance in MOOC for STEM. ICERI2020 Proceedings, 9788–9793.

Lin J (1991) Divergence measures based on the Shannon entropy. IEEE Trans Inf Theory 37(1):145–151

Mayring P (2010) Qualitative Inhaltsanalyse: Grundlagen und Techniken, 11th edn. Beltz

Miyake A, Kost-Smith LE, Finkelstein ND, Pollock SJ, Cohen GL, Ito TA (2010) Reducing the gender achievement gap in college science: a classroom study of values affirmation. Science 330(6008):1234–1237

Murphy C, Klotz AC, Kreiner GE (2017) Blue skies and black boxes: the promise (and practice) of grounded theory in human resource management research. Hum Resour Manag Rev 27(2):291–305

Myers MD, Newman M (2007) The qualitative interview in IS research: examining the craft. Inf Organ 17(1):2–26

Noeltner M, Kroenung J, Kuhlmann BG (2019) Disarming prejudice: how ease of use mitigates the detrimental effect of it-based stereotype threat on the it task performance of older adults.

Nouri J, Zhang L, Mannila L, Norén E (2020) Development of computational thinking, digital competence and 21st century skills when learning programming in K-9. Educ Inq 11(1):1–17. https://doi.org/10.1080/20004508.2019.1627844

Paloheimo A, Stenman J (2006) Gender, communication and comfort level in higher level computer science education-Case study. Proceedings. Frontiers in Education. 36th Annual Conference, 13–18.

Pennebaker JW (2011) The secret life of pronouns. New Sci 211(2828):42–45

Puyol-Antón E, Ruijsink B, Piechnik SK, Neubauer S, Petersen SE, Razavi R, King AP (2021) Fairness in cardiac MR image analysis: an investigation of bias due to data imbalance in deep learning based segmentation. Medical Image Computing and Computer Assisted Intervention–MICCAI 2021. In: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24, 413–423

Richard GT, Hoadley C (2015) Learning resilience in the face of bias: online gaming, protective communities and interest-driven digital learning. International Society of the Learning Sciences, Inc. [ISLS].

Rosenberg-Kima RB, Plant EA, Doerr CE, Baylor AL (2010) The influence of computer-based model's race and gender on female students' attitudes and beliefs towards engineering. J Eng Educ 99(1):35–44

Ruipérez-Valiente JA, Staubitz T, Jenner M, Halawa S, Zhang J, Despujol I, Maldonado-Mahauad J, Montoro G, Peffer M, Rohloff T, Lane J, Turro C, Li X, Pérez-Sanagustín M, Reich J (2022) Large scale analytics of global and regional MOOC providers: differences in learners' demographics, preferences, and perceptions. Comput Educ. https://doi.org/10.1016/j.compedu.2021.104426

Smith JL, Morgan CL, White PH (2005) Investigating a measure of computer technology domain identification: a tool for understanding gender differences and stereotypes. Educ Psychol Measur 65(2):336–355

Steele CM, Aronson J (1995) Stereotype threat and the intellectual test performance of African Americans. J Pers Soc Psychol 69(5):797

Strauss A, Corbin J (1990) Basics of qualitative research. Sage publications

Stumpf S, Peters A, Bardzell S, Burnett M, Busse D, Cauchard J, Churchill E (2020) Gender-inclusive HCI research and design a conceptual review. Found Trends® Hum-Comput Interact 13(1):1–69

Sultan U, Axell C, Hallström J (2020) Technical or not? Investigating the self-image of girls aged 9 to 12 when participating in primary technology education. Des Technol Educ: an Int J 25(2):175–191

Takács I (2022) Through a glass darkly: women in the scientific elite. Budapest: L'Harmattan Kiadó. https://doi.org/10.56037/978-2-14-027730-6

Vasilescu B, Capiluppi A, Serebrenik A (2014) Gender, representation and online participation: a quantitative study. Interact Comput 26(5):488–511

Vasilescu B, Capiluppi A, Serebrenik A (2012) Gender, representation and online participation: a quantitative study of stackoverflow. In: 2012 International Conference on Social Informatics, 332–338

Walton GM, Yeager DS (2020) Seed and soil: psychological affordances in contexts help to explain where wise interventions succeed or fail. Curr Dir Psychol Sci 29(3):219–226

Weck A (2022) Annahita Esmailzadeh: „Ich entspreche nicht dem Stereotyp eines IT-lers“. T3n. https://t3n.de/news/annahita-esmailzadeh-stereotype-it-tech-branche-1471342/

Willig C (2009) Introducing Qualitative Research in Psychology: Adventures in Theory and Method, 2nd edn. Open University Press, Maidenhead

Wrohlich K (2017) Gender pay gap innerhalb von Berufen variiert erheblich. DIW Wochenber 84(43):955–961

Yip Y-C, Yip K-H, Tsui W-K (2021) Exploring the gender-related perceptions of male nursing students in clinical placement in the Asian context: a qualitative Study. Nursing Reports 11(4):881–890

Ypma HV (2019) Eliminating Gender Inequity: Why Women Choose Computer Science. Northeastern University

Yücel Y, Rızvanoğlu K (2019) Battling gender stereotypes: a user study of a code-learning game, “Code Combat”, with middle school children. Comput Hum Behav 99:352–365

Zhou G, Ganesan K (2016) Linguistic understanding of complaints and praises in user reviews. Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp. 109–114

Authors and Affiliations

Daniel Stattkus[1] · Lorena Göritz[1] · Katharina-Maria Illgen[1] · Jan Heinrich Beinke[1] · Oliver Thomas[1,2]

✉ Daniel Stattkus
daniel.stattkus@dfki.de

Lorena Göritz
lorena.goeritz@dfki.de

Katharina-Maria Illgen
katharina.illgen@dfki.de

Jan Heinrich Beinke
jan.beinke@dfki.de

Oliver Thomas
oliver.thomas@dfki.de

1 Smart Enterprise Engineering, German Research Center for Artificial Intelligence, Hamburger Str. 24, 49084 Osnabrück, Germany

2 Information Management and Information Systems, Osnabrück University, Hambruger Str. 24, 49084 Osnabrück, Germany