A systematic content analysis of gender treatment in applied linguistics research

Filip Bigos

MSc in Applied Linguistics for Language Teaching, 2024

DECLARATION BY THE CANDIDATE AS AUTHOR OF THE DISSERTATION

1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.
2. I allow the Department to deposit on my behalf a copy of this dissertation in the Oxford University Research Archive ('ORA') where it shall be freely available online for use in accordance with ORA's Terms and Conditions of Use [https://ora.ox.ac.uk/terms_of_use].
3. I understand that this dissertation should not contain material that can be used to personally identify individuals or specific groups of individuals (unless permission has been obtained from the individuals) and that such material should be removed before this dissertation is deposited in ORA.
4. I agree to be bound by the terms of the ORA Grant of Non-exclusive Licence [https://ora.ox.ac.uk/deposit_agreements] and I warrant that to the best of my knowledge, making my thesis available on the internet will not infringe copyright or any other rights of any other person or party, nor contain defamatory material.
5. I agree that my dissertation shall be available for download in ORA in accordance with paragraphs 2, 3 and 4 above.

Signed [an electronic signature is sufficient]:	Filip Bigos
Date:	6 August 2024

A systematic content analysis of gender treatment in applied linguistics research

Dissertation submitted to the University of Oxford in partial fulfilment of the requirements for the degree of MSc in Applied Linguistics for Language Teaching

Filip Bigos

Word count: 19,989 words

Abstract

Gender, and *sex* are generally presumed to refer to two different concepts, namely socio-cultural expectations with regards to behaviour, and biology respectively. However, research on *gender* as an independent variable in applied linguistics research is often confounded with that on *sex*, and their lack of disambiguation leads to often conflicting, and inconclusive results in the field.

To substantiate this claim, this study aimed to critically investigate whether current applied linguistics research on gender includes the definition of the concept, whether it collects the information on participants' gender in a valid way, and to what extent its findings could be interpreted in line with the definition of gender. This definition presupposes that gender is a behaviour performed in an interaction, it is influenced by socio-cultural expectations, and it is context-dependent. Relevant research was sampled using a systematic review methodology to reduce bias, and to collect a representative sample of 98 studies, 36 of which met the inclusion criteria. Corpus analysis was conducted on said 36 studies to ascertain the presence of the definition of gender, and qualitative text analysis method was used to determine how the studies collected the information on the participants' gender. A narrower selection of 20 studies which reported a significant/ notable effect of gender was used for further qualitative text analysis, employing an inductive approach, to determine the extent to which the findings could have been interpreted in line with the definition of gender.

The study found that one study defined gender, five stated the method of collecting the information on gender, out of which one could have been potentially considered as valid. 50% of the studies could have been interpreted in line with the definition of gender. Such results point to an acute lack of validity,

and reliability in the examined research, and they suggest an explanation for the inconclusive state of gender research in applied linguistics.

To overcome this issue, this study calls for methodological innovation in the field. The term innovation is used because of the gravity of the issue: simple methodological rigour is insufficient. An increased methodological consciousness of the relevance of research methods used to investigate the concept of gender is believed to lead to more valid, more reliable, and hence more conclusive findings. In case of some research areas, such as language learning strategies (LLSs), this involves rethinking the concept to separate it into parts that are either behavioural/ cultural, i.e., those pertaining to *gender*, or cognitive/ biological, i.e., those pertaining to *sex*. In other areas, the validity, and reliability of research methodology ought to be increased by conceptualising gender, defining it, measuring it accurately, and ensuring that the methods used to investigate the dependent variables are reflective of the performative/ interactional, socio-cultural, and contextual nature of *gender*.

Dedicated to Dave Moreton.

You had a tremendous impact on the lives of so many, me included.

You may have left this world, but your love, legacy, and ever so optimistic spirit remain alive in us.

Oh, how great thou art!

Acknowledgments

First of all, I would like to acknowledge the support of my supervisor, Heath Rose, who believed in my idea right from the very start. Heath always encouraged me to continue exploring my ideas, reading, and pushing myself to persevere with this research. His research expertise was invaluable, as was his feedback on my work, and I would not have been able to complete this dissertation without his input – thank you!

I am also grateful to Faidra Faitaki who convened the individual and group differences module which inspired me to pursue this topic. Her passion for the topic was infectious, and her references were a useful starting point for this thesis. Further support came from Catherine Scutt, and other librarians who pointed me in the right direction, and taught me how to use databases. Thanks are extended to Olcay Sert, and Damon Young who shared some of their work, and ideas with me which informed different parts of my background research.

Professionally, I need to mention Sally German, Rachael Smith, and Stephanie Holme. These three teacher trainers are the source of my passion for teaching which drove me to undertake this master's degree, and I have to express my profound gratitude to them for inspiring me to be the professional I am today.

Special thanks are due to Benen Whitworth, my second reviewer, who dedicated her time to ensure methodological rigour in my research, as well as to Josh, Julie, Dave, Chris, Lily, Winny, and Dexter for listening to my ideas, sharing their views, and helping me find the right focus. Notably, the financial support from my brother, and my father was instrumental on this course, and I am profoundly thankful for them believing in me.

A very special mention needs to be given to my partner, Zach, who has been an absolute rock on which this work was built. He not only supported me, and encouraged me to pursue my passion, but he also ensured I did not have to worry about mundane chores, or anything that could have distracted me from this endeavour. He was the one who pushed me to focus, and made me prioritise studies over other things in life when deadlines were near. Words cannot express how much you being there meant to me over the course of this work.

Table of Contents

List of tables and figures

List of acronyms and abbreviations

ADHD	Attention deficit hyperactivity disorder
ASRS	ADHD Self-Report Scale
ELT	English language teaching
EFL	English as a foreign language
ESL	English as a second language
FCE	First Certificate in English
FL	Foreign language
GPA	Grade point average
KWIC	Key Word in Context
LGBTQ	Lesbian, gay, bisexual, transgender, and queer
LLS	Language learning strategy
LMOOC	Language Massive Open Online Course
L2	Second language
NCVS	National Crime Victimisation Survey
PICOC	Population, intervention, comparison, outcomes, and context
PPVT	Peabody Picture Vocabulary Test IV
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
RQ	Research question
SILL	Strategy Inventory for Language Learning
SOLO	Search Oxford Libraries Online
UNESCO	United Nations Educational, Scientific and Cultural Organization
The UK	The United Kingdom
The US	The United States (of America)

1. Introduction

1.1 Background and rationale of the study

Gender has been a long-acknowledged group difference in language learning, and a considerable body of research has explored whether, and to what extent, gender has an impact on language acquisition. In applied linguistics, this research has included studies examining the effect of gender on academic attainment, performance in language skills, classroom behaviour, or the use of language learning strategies (LLSs), to name a few.

However, the existing research on gender as an independent variable in applied linguistics is often inconclusive, and studies frequently produce findings which are contradictory to prior research. Whilst the existence of conflicting evidence in research is not in itself surprising—and it could be argued that it is an essential part of searching for knowledge—such profound inclusiveness as is found in the research on gender in applied linguistics indicates a potential presence of deep-rooted issues in the existing body of research.

The issues can start to be observed by looking at research examining the impact of gender on behaviour: for example, Ogbay (1999) found that females were more reluctant to speak in a classroom, and Chavez (2000) reported that in higher education, females were more likely to want to please the teacher with the accuracy of their input. It is important to note that behaviour is cultural, and gender, i.e., how females and males are expected to behave, is a cultural concept. Ekstrand (1980) supports this notion when claiming that behavioural differences are likely to be attributable to cultural factors, rather than cognitive ones. As a result, it can be posited that such findings should be assigned to *gender*, and due diligence should be paid to differentiating this from *sex*, which is a biological factor. By extension, understanding the disambiguation of these

two concepts ought to be fundamentally reflected in the methodology of studies on gender regardless of what dependent variable they are researching.

This issue is important because researchers should aspire to be methodologically rigorous in their work, and a study design can have a significant impact on what findings the study generates. It is, therefore, posited that methodological rigour, or lack thereof, has had an impact on the conclusiveness of applied linguistics research pertaining to gender.

1.2 Aims of the study

This dissertation adopts a critical approach to this issue, and it aims to collect a systematic sample of current applied linguistics research on gender to explore whether the research is methodologically rigorous enough to generate findings that pertain to *gender*[1], or whether these could be assigned to a different variable, such as *sex,* or cognition. To do this, this study posits three specific research questions (RQs):

> RQ1: Does current research examining gender in applied linguistics include its definition?
>
> RQ2: Does current research examining gender in applied linguistics include a valid method in which the information on the participants' gender was collected?
>
> RQ3: To what extent can the findings of current research examining gender in applied linguistics be interpreted in line with the definition of *gender*?

[1] As the object of the study is the term *gender*, it is often italicised when a particular distinction needs to be made between it, and biological *sex* which would be italicised for the same reason.

1.3 Dissertation outline

This dissertation starts by defining the concept of gender by drawing on feminist studies, before providing an overview of the existing gender research in applied linguistics, highlighting its inconclusive nature. The last part of chapter two reviews how gender has been methodologically approached in wider social science research. Chapter three details the systematic approach used to collect studies for this dissertation, as well as the approach to their analysis which is reported in chapter four. This results chapter is organised by research questions, and, drawing on methodological rigour, its findings are discussed in section five. The findings of the study, its limitations, and recommendations for future research are made in chapter six.

2. Literature review

2.1 Defining gender

Although the distinction between *gender* and *sex* has been contested since the inception of the former, nowadays they are universally accepted as denoting two different concepts. *Sex* refers to a set of biological factors pertaining to a human being (chromosomes, physical features, sexual organs, etc.), whereas *gender* encompasses social factors (behaviour, culture, role in a society, etc.). This can be observed in the definitions provided by English dictionaries. The definition of *sex* reads as follows: "the two sexes are the two groups, male and female, into which people and animals are divided according to the function they have in producing young" (Colllins, n.d.). In contrast, *gender* is defined as "a group of people in a society who share particular qualities or ways of behaving which that society associates with being male, female, or another identity" (Cambridge University Press and Assessment, n.d.). A specific distinction between the two is made in the following definition of *gender*: "the fact of being male or female, especially when considered with reference to social and cultural differences, rather than differences in biology; members of a particular gender as a group" (Oxford University Press, n.d.). As dictionaries reflect the common usage of language by its users, it is observed that *sex* and *gender* are generally understood to mean two different concepts. This thesis explores whether the two are sufficiently differentiated in current applied linguistics research.

2.1.1 Rise in feminism

The need to differentiate *gender* from *sex* arose during the second wave of feminism in the 1960s. Prior to that, a biological determinist view, which argued that behavioural and psychological traits are a result of one's biology (Geddes, 1895), was used to discriminate against women, as the position enabled it to be claimed that since the differences between the sexes were rooted in biology, they were unchangeable (Nicholson, 1994). This debate has not completely

ceased to exist – as late as the 1990s, Gorman and Nash (1992) posited a number of claims highlighting how women's different biology affects their mental processes, with Rogers (2000) taking this even further by arguing that women should be precluded from certain occupations because of their hormonal instability.

2.1.2 Nature vs nurture

The biological distinction between women and men certainly needs to be acknowledged. If it did not exist, women and men would compete together at sporting events. Even the International Olympics Committee regulates biological factors which make transsexual athletes eligible to compete as their new sex (International Olympic Committee, 2004). Equally, hormones and emotions, both physiological processes, affect our behaviour, making biology central to who we are as human beings.

To complicate the gender-sex dichotomy further, some feminist theorists note the importance of sex in determining gender. In her "coat rack" theory, Nicholson (1994) posits *sex* as "the site upon which gender was thought to be constructed" (p. 81). This means that societal expectations with regards to women's, and men's behaviour, and their conformity to the societal "norms" are based on the sex of the people. This is, ultimately, where gender stereotypes originate, and why many tend to think of *gender*, and *sex* as synonymous. This is further confounded by arguments that women's, and men's physical features can be affected by social practices. Fausto-Sterling (1992) contends that body dysmorphia is a result of unequal exercise opportunities for people of different sexes in certain societies. Similarly, Jaggar (1983) notes that women being given less food as a result of societal hierarchy can not only lead to malnutrition, it also makes their bodies physically smaller, something which is also evident in practices, such as Chinese traditions of foot binding. Such effects of *gender* (societal factors) on *sex* (biological factors) tangle the distinction further, and

they might suggest that separating *gender* from *sex* is futile, as the two are so intertwined.

However, distinguishing them is not only important to the advancement of the society, but the biological determinist views undermining the value of *gender* as a separate concept have also been called into question. Stoller (1968) was amongst the first to use the term *gender* to refer to one's levels of femininity/ masculinity – separating them has enabled the two terms to disintegrate, thus allowing female bodies to be gendered as male, or vice versa (Haslanger, 2000; Stoljar, 1995). This is useful not only to better understand transsexual, and transgender people whose biological sex might not align with their felt gender, but it was also essential to progress women's rights. Rubin's (1975) positioning of *gender* as "socially imposed division of sexes" (p. 179) enabled the feminist movement to argue that such differences between females and males were possible to change. Like with the argument of emotions and hormones affecting our behaviour, albeit hard, they are susceptible to be controlled by the mind. Change over time in biological factors was also one of the primary pillars of Fausto-Sterling's (2000) refutation of the biological determinists views presented by Gorman and Nash (1992): corpus callosums, which was central to their argument of women's lack of biological predisposition for certain tasks, has been shown to be the same size in infants of both sexes, thus giving rise to the notion that the change in its size might be due to nurture. Therefore, it could be argued that even the physical attributes commonly associated with biological *sex* can still be questioned.

2.1.3 Gender as a performative act

The exigency to accept gender as a standalone product of culture, and society also lies in the variety in which culture is expressed. One way to approach this is through Butler's (2006) performative perspective of *gender*. They question whether *gender* is something one "has", or "is", and if it is constructed, by whom.

Butler (2006) posits that *gender* is something people "do" through both language use, and repetitive actions every day. People's ability to construct gender, according to Butler (2006), enables them to choose their gender, thus diminishing the importance of the sexed body, and making the distinction between *sex*, and *gender* unintelligible. Whilst Butler (2004) might be right in that the performativity of *gender* enables one to deconstruct the norms which attempt to define genders, their theoretical analysis does not abolish the very existence of these norms. In reality, each society has its own norms with regards to what it means to be "female", or "male". The commonality problem within feminist theories highlights the complexity of this – if you compare women from different societies, times, of different races, and social backgrounds, they might have very little in-common apart from their biology (Spelman, 1988). Such socio-cultural differences also hold true for men, and are often cited as stereotypes: the Italians are said to be passionate, the Germans effective, the Japanese polite.

2.1.4 Gender as a socio-cultural construct

All of this leads this dissertation to posit that *gender* is an expression of socio-cultural expectations enforced onto people of particular *sex*. These expectations vary by culture and society, even one's position within it. *Sex* is viewed to be "between the legs", and *gender* "between the ears" with autonomy, and agency over our actions being intrinsic to the disambiguation. It is people's decision making, and the ability to change and control their circumstances, behaviour, and self-presentation, however hampered by societal expectations they might be, that distinguish us from animals which do not have genders. Whilst sexed bodies live in a society, they will be included in, and influenced by culture, yet their biology will remain intrinsically different from their cultural expression. The discussion about the extent to which the biological, and the cultural are distinct is, in the context of social sciences, and this thesis, instrumental, since being

able to draw a line between *sex,* and *gender* enables us to attribute research findings to either biology, or culture respectively.

2.2 Gender in applied linguistics research

Gender has been a subject of multiple studies within applied linguistics. The research within the field can be broadly categorised into two sub-categories – research exploring how gender is represented within English language teaching (ELT) materials, and research investigating gender as an independent variable in research populations of learners, or teachers.

2.2.1 Representation of gender in ELT materials

A UNESCO report (Blumberg, 2008), examining gender bias in course books, points at evident gender bias towards males which has been corroborated by the body of research on this topic. In texts, recordings, and visuals, males tend to be represented more frequently (Ansary & Babaii, 2003; Aydınoğlu, 2014; Barton & Sakwa, 2012; Lee, 2014), and there is a notable trend to use gender terms, such as *a fireman* instead of *a firefighter*, as well as the male pronouns *he/ him/ his* to refer to a character whose gender is not known (Barton & Sakwa, 2012; Lee, 2014).

Apart from gender bias, ELT materials also tend to perpetuate gender stereotypes. Males tend to have a wider range of occupations, whilst females usually work in gender-stereotyped roles, such as teachers, or nurses (Ansary & Babaii, 2003; Otlowski, 2003). Males have been found to be presented as more active, often being outdoors, and engaging in activities such as playing sports, or working on cars, whilst females tend to be more passive, and mostly indoors, e.g., doing housework, or reading (Ansary & Babaii, 2003; Aydınoğlu, 2014; Barton & Sakwa, 2012; Jasmani et al., 2011; Lee, 2014; Otlowski, 2003). Stereotypical are also the descriptions of females which tend to fixate on

appearance, emotions, or age (Barton & Sakwa, 2012; Lee, 2014; Şeker & Dinçer, 2014; Söylemez, 2010).

Recently, newer course books have started representing genders more fairly (Gray, 2000; Hilliard, 2014; Lewandowski, 2014; Mineshima, 2008; Yang, 2012). Yilmaz (2012) found that the newest versions of course books *Total English, Cutting Edge,* and *New Headway* had nearly equal ratios of male, and female visibility, and there had been a notable improvement in eradicating some gender stereotypes, as well as gender-biased language since the books' first editions. Improvements have also been noted in Business English, and English for Specific Purposes materials (Adel & Enayat, 2016; Goyal & Rose, 2020), however, the progress in these is hampered by the existence of gender bias in the contexts which the materials are representing.

Research has also been conducted on teacher training materials where gender stereotypes were found to be present, too (Zittleman & Sadker, 2002), as well as on how teachers respond to, and use teaching materials to promote gender equality. The latter was examined by Barton and Sakwa (2012) who quoted their Ugandan teacher participants as saying that they did not want to express their own opinions in front of students, or that the gender bias in course books was an apt reflection of their culture.

In summary, despite recent improvements, gender bias, and gender stereotypes are evident in ELT teaching materials. This can possibly be a reflection of the culture where the materials were published, or the context which they represent. Equal attention should be paid to teacher training to ensure teachers are equipped to deal with issues arising from the situation.

2.2.2 Gender as an independent variable

Applied linguistics research also investigates gender as an independent variable in relation to dependent variables, particularly the effect gender has on language use, performance in the four language skills (speaking, writing, listening, reading), learning outcomes, attitudes, motivation, and LLS use.

2.2.2.1 Gender differences in language use

A meta-analysis of studies on talkativeness, affiliative, and assertive speech by Leaper and Ayres (2007) has found men to be more talkative than women (d = -0.14), that women used affiliative speech more (d = 0.12), and men used more assertive speech (d = 0.09). Although these findings were statistically significant, and the analysis of the effects of a number of variables informative, the effect sizes were not big. The findings with regards to talkativeness were supported by Hall (1984) and James and Drakich (1993), but refuted by others in different contexts (Hyde & Linn, 1988; Leaper et al., 1998; Leaper & Smith, 2004). Such inconclusiveness of this body of research might be a result of lack of clarity pertaining the attributability of talkativeness to either biological sex differences, or gendered performance based on socio-cultural expectations.

2.2.2.2 Gender differences in language skills

Related to language use are also speaking skills which were explored by Ogbay (1999) who found that females were more reluctant to speak in class. Females fared significantly better than males in standardised reading tests in a longitudinal study of 7,075 school-age participants by Robinson and Lubienski (2011). Female superiority, however, does not hold true for other skills. Morris (1998) investigated the quality of writing produced by 42 ESL students. The study found that although females scored higher marks, which was a result of their better adherence to writing guidelines than that of males, the writing of both genders was linguistically comparable. Similarly, no significant gender

differences were observed by Bacon (1992) in her study of listening comprehension of 50 learners of Spanish as a foreign language (FL). On the contrary, males scored significantly higher in listening vocabulary tests (p = 0.000, and p = 0.0033 in the two respective listening tests) than females in Boyle's (1987) study of 285 ESL students at a Chinese university. What is unclear from this body of research is whether these findings can be attributed to gender – one could hypothesise that reluctance to speak in class, or spending more time studying could be a result of socio-cultural expectations, yet the same would be more difficult to say for performance in receptive language skills which are performed individually, and internally, thus being more likely subject to biological/ cognitive influences, rather than cultural ones.

2.2.2.3 Gender differences in learning outcomes

Boyle's (1987) study also included 10 language proficiency tests in which females scored higher than males, with results of seven of the tests reaching statistical significance. Females (n = 43) also outperformed males (n = 74) in a number of comprehension, immediate vocabulary, and vocabulary retention tests in Lin's (2011) study investigating the gender differences on a video-based assisted language learning course. A study on a bigger scale was conducted by Van Der Slik et al. (2015) who explored the effects of gender on L2 acquisition, and on the L2 Dutch proficiency of 27,119 immigrants, as measured by the State Examination of Dutch as a Second Language. Females outperformed males in speaking ($p < 0.001$), and writing ($p < 0.001$), males did better in reading (p = 0.002), but no significant differences were found in listening proficiency (p = 0.232). Thus, it should be problematised whether such results can be blankly assigned to *gender*, or whether they need to be assessed more carefully, and disambiguated based on whether they are influenced by biological, or socio-cultural factors.

2.2.2.4 Gender differences in attitudes, motivation, and behaviour

Applied linguistics research has also investigated gender differences with regards to attitudes, motivation, and behaviour of FL learners. Taylor and Marsden (2014) investigated the attitudes of 604 secondary school pupils towards learning an FL, which were found to be more negative in boys than in girls. Such results could be linked to motivation: Bacon and Finnemann (1992), in their study examining the relationship between gender, and the beliefs of 938 L2 learners of Spanish found females to have significantly higher levels of instrumental motivation. This was corroborated by Koul et al. (2009) who, focusing on measures of goal orientation, also found their female participants to have significantly lower levels of socio-cultural goal orientation than males ($p < 0.01$). However, conversely to these two studies, Ludwig (1983) found male FL students of French, and German to be more instrumentally motivated than females. A different study by Chavez (2000) explored how males, and females behave differently in the classroom based on the gender of their classmates, and the teaching assistants. Analysed using a 100-item questionnaire, the results from 201 German FL students included a lot of interesting findings, from females being more self-conscious, males not wanting to work with other males, to females thinking that males did not choose the topics of the lessons often enough. Such findings could be broadly interpreted as being subject to socio-cultural expectations, as the different results reported by the researchers could be a result of their context.

2.2.2.5 Gender differences in language learning strategy use

The earlier-mentioned study by Bacon (1992) also explored the use of language learning strategies (LLSs) by people of different genders. The study found that females reported using significantly more metacognitive strategies ($p < 0.02$), whilst they were more consistent in their use of bottom-up strategies when listening than males. Similar findings were noted by both Oxford and Nyikos

(1989), and Ehrman and Oxford (1989) whose female participants reported using LLSs more often than males. More detailed were the results of a study by Liyanage and Bartlett (2012) who surveyed 886 Sri Lankan ESL (English as a second language) learners. Although they support the findings that females, on the whole, report using LLSs more often than males, males reported using *self-monitoring*, *translation*, *repeating*, and *asking questions for clarification* more often than females. Young and Oxford (1997), in their study involving 49 FL learners of Spanish, also found males to favour three specific strategies statistically significantly more than females, and females reported using two strategies significantly more. However, on the whole, this study did not find significant differences between the strategies used by participants of different genders.

2.2.2.6 Summary of gender differences

In summary, the research on gender differences in applied linguistics seems to be inconclusive. Studies have either found no significant effect of gender in their studies, or if they have, other studies seem to have refuted their findings. Whilst these results could be attributed to the limitations of individual studies, this dissertation posits that the underlying reason for the inconclusiveness in the field is the lack of methodological rigour pertaining to differentiating *gender*, and *sex* as two different concepts. This could be observed in a more detailed analysis of the above-mentioned results of Liyanage and Bartlett's (2012) study: although females reported more frequent use of cognitive, and metacognitive LLSs, the use of social-affective strategies was more mixed depending on the participants' ethnic group. This suggests that the use of socio-affective strategies might be subject to socio-cultural expectations with regards to interacting with other people in the different ethnic groups, i.e., it is more context, and culture-dependent, rather than being attributable to the biological sex of the participants. On the other hand, the use of cognitive, and

metacognitive LLSs might be more cognitively/ biologically driven, thus being likely to be attributable to *sex*, rather than *gender* – a distinction that should be made in the methodology of the studies. Although certain researchers, such as Leaper and Ayres (2007), or Van Der Slik et al. (2015) mention different interpretations of their data, such as nature-, or nurture-based theories, the fundamental lack of their differentiation from the onset of the studies is worthy of critical examination.

2.3 Methodological approaches to gender in social science research

In the broader field of social sciences, which applied linguistics is a part of, information on gender started being collected for empirical research purposes in the 1970s to investigate it as a variable (Stacey & Thorne, 1985). By the 1990s, gender became widely researched in many disciplines, particularly in anthropology, sociology, and history (Curthoys, 2014), and since then it has gained momentum. Now, data on gender is collected to promote gender equality, especially when it comes to access to health services, to monitor discrimination in employment, or political participation, and gender representation in media (Curthoys, 2014; Magliozzi et al., 2016), to name just a few applications.

2.3.1 Conventional methods

Traditionally, information on gender tends to be measured on a dichotomous (female – male) scale where participants can choose from two options. As such, the measure tends to be "taken for granted" (Magliozzi et al., 2016, p. 2) with insufficient consideration being given as to how this information should be collected. This stems from the "theorisation of heteronormativity – the suite of cultural, legal, and institutional practices that maintain normative assumptions that there are two, and only two genders, [and] that gender reflects biological sex" (Kitzinger, 2005, as cited in Schilt & Westbrook, 2009, p. 441). The resistance to measuring gender beyond this dichotomous gender binary is

systematic (Tabler et al., 2023), because people presume that appearances are gendered, and based on the biological reality of one's sex (West & Zimmerman, 1987), as alluded to in the "coat rack" theory (Nicholson, 1994).

2.3.2 The two-question method

These perceptions have been disrupted by the increasing visibility of transgender people whose gender identity is not aligned with the biological sex assigned at birth. The rise in trans activism has initiated a shift in research methodology towards using a two-question approach whereby participants are asked about their present gender identity, and their sex assigned at birth separately (GenIUSS, 2014). This method was used by Truman et al. (2019) who investigated its inclusion in the NCVS (National Crime Victimisation Survey used by the United States Census Bureau), focusing on interviewers', and respondents' reactions to questions on sexual orientation, and gender identity. Using a mixed-method approach, the researchers analysed 899 responses to a debriefing questionnaire, together with qualitative data from focus groups, and interviews with interviewers who had administered the NCVS to respondents. Focusing on the gender identity questions, the results indicated that, on the whole, respondents reacted positively to the questions, particularly people who identify as LGBTQ, with 52% of interviewers reporting no issues. 39% of them reported a negative reaction, with some respondents expressing heteronormative views, many questioning the need to include the questions, and the elderly not understanding the terminology. Such qualitative data provides an interesting insight into people's perception of the need to include more inclusive gender identity questions in research. Having said that, the validity of the quantitative data needs to be questioned – since it was obtained from the interviewers, not interviewees, it is unclear whether the 39% of interviewers had issues with one respondent each, or whether the reported negative reactions came from multiple interviewees. The non-response rate of the gender identity question

was low (0.97%), and mostly related to the interview mode, and age.

Furthermore:

> Among transgender respondents, 51.7% identified as transgender on the current gender identity question and 48.3% reported discordant sex at birth and current gender identity. These data indicate that it is important to collect gender identity using the two-step method to provide an accurate measure of the transgender population. (Truman et al., 2019, p. 847)

These findings indicate that whilst some respondents might question the need to be asked about their gender identity, the information is provided. Pertinently to this paper, it is a method which attempts to separate biological *sex* from *gender*. In the field of feminist studies, it is subject to critique, since it reduces gender, a socially-constructed concept, to a still dichotomous variable without being able to capture the nuances of gender identity (Stacey & Thorne, 1985).

2.3.3 The sliding scale method

This criticism has led to the proposal to use sliding scales to gather information on participants' gender identity, a method which was employed by Magliozzi et al. (2016). The researchers designed a survey which asked participants not only about their sex at birth, and gender identity through categorical, multiple-choice options, but also about their own perception of themselves, and others' perceptions of them on a seven-point Likert scale, measuring both how feminine, and how masculine they were. The survey had been piloted, and then administered to adult US residents online using Amazon Mechanical Turk, generating 1,522 responses. The results indicated that although 99% of participants were cisgender according to their categorical answers, only 24% chose the polarisation score of seven on the Likert scales. This implies that the remaining male participants did not see themselves as "very" masculine, and females as "very" feminine, meaning their masculinity, and femininity scores overlapped. 7% of respondents gave themselves identical masculine, and

feminine scores, and further 4% scored less on the scale that corresponds to their sex assigned at birth. 209 participants, who provided open-ended feedback, indicated that they had considered their personality traits, job, pastimes, and appearance when assessing their masculinity – femininity. Whilst these might be perpetuating gender stereotypes, the scales captured a more nuanced picture of gender identity than conventional methods.

2.3.4 The importance of methodological rigour pertaining to gender

Having reviewed three methodological approaches to measuring gender, a question of how important this is to researchers arises. This was examined by Tabler et al. (2023) who, using a 32-question questionnaire, surveyed the attitudes of 309 faculty members at US universities towards expansive gender, and sexuality measurements in general, and to their own research. The results indicated that although 65.7% indicated that expansive gender identity measurements were either "extremely important", or "very important" in general, only 36.6% said the same in relation to their own research (the findings pertaining to sexuality measures were similar). This finding is notable in that despite academics' beliefs that expansive measures of gender identity are important, they do not consider them so in relation to their research. This disparity could be caused by the participants' respective research topics, where gender might not be relevant, which a further qualitative element of the study could have ascertained. Illuminating was also the finding that participants who identified as LGBTQ found expansive gender measurements more important, however, quantitative researchers, and those in teaching positions found them less valuable. The authors speculate that these, respectively, could be the effects of the limited scope of teaching-focused staff to conduct their own research, and the contribution of qualitative research to the creation of expansive measurements of gender, thus making quantitative researchers less understanding of the issues.

It needs to be acknowledged that some researchers might be marginalised for pursuing gender expansive topics, and others, in fields such as clinical psychology, might tend to ignore gender diversity completely (Tabler et al., 2023). Regardless of the study's limitations, including its low response rate of less than 20%, its findings indicate that "despite framings of academic contexts as liberalizing and faculty valuing evidence-based, inclusive viewpoints" (Gross & Simmons, 2014, as cited in Tabler et al., 2023, p. 15), expansive measurements of gender identity are not of high importance to many researchers.

2.4 Summary, and research questions

2.4.1 Methodological rigour pertaining to gender in applied linguistics research

As outlined in this chapter, *gender*, and *sex* are two different concepts, referring to one's culture, and biology respectively. The two terms are representative of nurture-, and nature-based theories, which some applied linguists have drawn on to discuss their findings. Despite the fact that the current methodology of collecting data on gender stems from the need to recognise trans people, and to better capture a variety of genders, the fundamental distinction between *sex,* and *gender* needs to be recognised in applied linguistics research, too.

A glance at the definition of *gender* in the *Longman dictionary of language teaching and applied linguistics* suggests that sufficiently differentiating it from *sex* has not been a priority in applied linguistics research, since the dictionary posits that *gender* "refers to sex as either a biological or socially constructed category" (Richards, 2013, p. 240). Although the definition conflates the two concepts, the 4th edition of the dictionary also acknowledges Butler's (2006)

view of gender as a performative act, suggesting that the views on this might be beginning to shift.

More accurate, and expansive measurements of gender might be vital in terms of promoting inclusiveness of underrepresented populations in research, and participants have been noted to prefer them (Tabler et al., 2023). Methods, such as sliding scales, which are used to this effect, can also capture shades of gender for people who might not identify as completely female, nor male, but have an identity somewhere in between. As useful as these might be in terms of promoting inclusivity, and our understanding of the human nature, this paper argues that the importance of gender measurements needs to be seen in their promotion of methodological rigour, and research validity. As Leaper and Ayres (2007) note, "students in research methods classes commonly learn that how a construct is measured can affect the particular results one finds" (p. 331). It could be hypothesised that asking participants whether their *gender* matches their *sex* assigned at birth might not yield a statistically significant effect on the results of a given study. However, not asking participants this question, or not even allowing them to self-identify their own gender points to a lack of awareness on the side of the researchers of what is that they are trying to explore – biology (sex), or culture (gender), thus resulting in contradictory research outcomes in the field.

2.4.2 Research questions

This paper aims to systematically analyse how *gender* is treated in applied linguistics research, and whether it is being sufficiently distinguished from *sex*.

To explore this, the following research questions (RQs) are posited:

> RQ1: Does current research examining gender in applied linguistics include its definition?

RQ2: Does current research examining gender in applied linguistics include a valid method in which the information on the participants' gender was collected?

RQ3: To what extent can the findings of current research examining gender in applied linguistics be interpreted in line with the definition of *gender*?

3. Methodology

3.1 Introduction

This mixed-method study is systematic in that it applied a systematic review methodology to compile, code and evaluate studies in an unbiased, replicable way. The study itself is not a systematic review, because its aims are not to synthesise a comprehensive body of research on gender. Rather it adopts methodological elements of the approach to ensure the collection of samples of research were carried out in a manner that reduced researcher bias.

Once compiled, the sampled studies were then used to create a corpus which was analysed using corpus linguistics techniques to answer RQ1, since corpus linguistics techniques "explore patterns of language use within text"(Wang, 2020). RQ2, and RQ3 were answered using qualitative text analysis methodology.

This chapter provides the methodological detail with regards to how the reviewed studies were complied, screened, and analysed, so that the research can be replicated. As this research concerns itself with methodological rigour, its systematic approach is crucial, since understanding past research methodology is needed to advance new methodological practices (Petticrew & Roberts, 2006).

3.2 Systematic collection of studies

3.2.1 Conceptual framework

Designing a conceptual framework is an essential step of a systematic approach (Newman & Gough, 2020), since it informs the research methods used to answer the RQs (Petticrew & Roberts, 2006). Further framing these as hypotheses can enable systematic reviews (or studies using systematic approaches) to test

these hypotheses, thus allowing the RQs to be clearly answered (Mulrow, 1994, as cited in Petticrew & Roberts, 2006).

Based on the information contained in the literature review, the following hypotheses can be made:

> Hypothesis 1: Most current research in applied linguistics investigating the concept and/ or impact of gender does not include a definition of the concept.
>
> Hypothesis 2: Most current research in applied linguistics investigating the concept and/ or impact of gender either does not stipulate how the information on participants' gender was collected, or it is determined based on the perceived participants' gender identity, i.e., external characteristics, such as clothing, or personal style, in line with the "coat rack" theory, and heteronormative views.
>
> Hypothesis 3: Since methodological shortcomings pertaining to the lack of due diligence of carefully defining the concepts of *gender,* and *sex* are presumed to exist, it is hypothesised that the findings of applied linguistics studies examining gender might not be able to be interpreted in light of how gender is defined.
>
> Hypothesis 4: As a further result of methodological shortcomings, it is hypothesised that most current research in applied linguistics investigating the concept and/ or impact of gender aligns participants' *gender* with *sex,* based on heteronormative views.

As the research questions concern themselves with methodological rigour of existing research, not with their participants, their review started broad. This was to ensure no important studies were missed (Macaro, 2020), that the findings of this study can be generalised to research conducted in multiple settings (Petticrew & Roberts, 2006), thus providing a fair overview of the state of applied linguistics research with regards to its treatment of gender. A narrower approach

was used to answer RQ3 to be able to dedicate the text analysis sufficient attention (Macaro, 2020), and clearly defined criteria were used to select the most relevant studies for this level of analysis.

3.2.2 Protocol

Petticrew and Roberts (2006) highlight the need to establish whether a systematic review on the research topic has already been conducted. In case of this paper, the International Database of Education Systematic Reviews (Chalmers, 2023), which is relevant to the field of applied linguistics, was searched for the term *gender* on 20 December 2023, and no relevant studies were retrieved. The systematic methods used to compile the studies are usually summarised in the protocol document (Newman & Gough, 2020) – because this study is not a systematic review as such, this methodology section is viewed as a sufficient representation of the protocol document.

3.2.3 Inclusion/ exclusion criteria

To better frame the focus of the RQs, Petticrew and Roberts (2006) propose using the PICOC model, where the acronym stands for population, intervention, comparison, outcomes, and context. With regards to population, this research concerns itself with studies conducted on adult learners of foreign languages. This is because according to existing data (Sun et al., 2022), gender dysphoria (i.e., when one's gender identity is not aligned with their sex at birth) peaks at the ages of 19 for natal women, and 23 for natal men, thus suggesting that the differences between *sex*, and *gender* become more relevant, and pertinent in adulthood.

Intervention, comparison, and outcomes are of less importance to this study since no data/ results of the compiled studies were synthesised. The context of relevant studies was restricted to those concerning second language acquisition of foreign/ second languages, since this study examined studies in

the field of applied linguistics, and to those conducted after 2004 when the Gender Recognition Act 2004 was passed in the UK. Although this research was not solely focused on studies carried out in the UK, and no limitations with regards to the location where the studies had been conducted were imposed to ensure sufficient breadth of studies, the time period covered in a systematic study needs to restricted (Petticrew & Roberts, 2006). Badger et al. (2000) illustrate choosing the publication of a UK legislation as the start date of their systematic review. The Gender Recognition Act 2004 legally recognised transsexualism, acknowledging the difference between *sex,* and *gender* in the process (Sandland, 2005), thus offering 2004 as a date when the disambiguation between the two terms came into the public domain in the UK. Due to the UK's influential position in social sciences, and human rights, its legislation could be viewed as potentially trend-setting. It could also be viewed as a point in time when many governments in the Western world were reviewing their laws and policies in relation to gender, based on research emerging at the time.

Although other criteria, such as the language of the study, or participant characteristics are commonly used in systematic reviews (Newman & Gough, 2020), these were of little interest in this study. As this study aims to evaluate the treatment of gender in current applied linguistics research, and it does not aim to synthesise results of studies, only a representative sample of studies is required. For these purposes, only studies written in English, published in an academic journal, and with a full reference to enable traceability were sought. Thus, the final sample can be said to be representative of research published in English language, indexed applied linguistics journals, rather than all existing research in the field. All types of studies were included, including qualitative, quantitative, and those not containing empirical data. Although Macaro (2020) warns that the exclusion of empirical studies may lead to synthesising theories, and opinions, due to this study's focus on their definitions and methodology,

and interpretation of gender, all types of studies were included. The full inclusion/ exclusion criteria used can be seen in Table 1.

	Inclusion criterion	**Exclusion criterion**
C1	Focused on modern language learners/ teachers	Not focused on modern language learners/ teachers
C2	Published after 2004	Published before 2004
C3	Published in an academic journal	Not published in an academic journal
C4	The object of the study are human participants	The object of the study are not human participants
C5	Includes adult participants over the age of 18	Does not include adult participants over the age of 18
C6	Complete reference	Incomplete reference
C7	Gender is explicitly addressed as a participant variable	Gender is not explicitly addressed as a participant variable
C8	Written in English	Written in languages other than English

Table 1: Inclusion and exclusion criteria

3.2.4 Search strategy

The conceptual framework and eligibility criteria inform the search strategy (Newman & Gough, 2020). As this study employed a systematic review methodology only to identify the studies to be analysed, and it is not a systematic review in itself, reliance on a limited number of electronic databases was warranted. Because of this, neither grey literature nor citation searching were conducted. The Scopus database had been chosen for its comprehensive coverage of applied linguistics research, and it was accessed electronically using institutional access through the Bodleian library. Such decisions are in line

with previous research that have collected journal articles as source data for analysis of applied linguistics topics (see, for example, Thomas et al. (2021)).

A search strategy needs to be sensitive (i.e., returning all potentially relevant studies), and specific (i.e., excluding irrelevant studies) (Petticrew & Roberts, 2006). The inclusion/ exclusion criteria were used to create a search string which included multiple synonyms to ensure a balance between sensitivity and specificity (Petticrew & Roberts, 2006).

The search string was pilot-tested (Victor, 2008). As the term “gender” needed to be explicitly addressed, it had to be included in the title, abstract, or keywords. Same criterion was used for the terms “adult*”, “English as a foreign language”, “English as a second language”, their acronyms (EFL and ESL respectively), and “modern language*”. An asterisk was used to include different morphological variations of the terms. In the pilot-testing, the term “L2” was added to the string which returned 444 results. As this was considered too many to screen for a study of this scope, this term was dropped. The final search string used can be found in Figure 1.

```
TITLE-ABS-KEY ( gender ) AND TITLE-ABS-KEY ( "English as a foreign language" OR "English as a second language" OR esl OR efl OR "modern language*" AND "adult*" ) AND PUBYEAR > 2003 AND PUBYEAR < 2025 AND ( LIMIT-TO ( DOCTYPE , "ar" ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) )
```

Figure 1: Search string

3.2.5 Screening, and data extraction

As the first trawl can produce too many studies to analyse in detail, their titles, abstracts and keywords need to be screened to ensure the studies’ relevance to the RQs (Macaro, 2020), that they meet the inclusion criteria, and to eliminate duplicates. As this is best done using two reviewers (Petticrew & Roberts, 2006),

a fellow student from the same cohort of this post-graduate programme as the author of this study was recruited. This reflects best practice in second language acquisition research, since the second screener came from the same field (Macaro, 2020), and had similar levels of experience. The reviewer was briefed on the aims of the study, its research questions, the methodological approach of the study, and the inclusion/ exclusion criteria, which were all provided both in writing, as well as discussed in a Teams meeting to ensure the second reviewer was clear about their role. To be methodologically robust, the second reviewer was asked to screen 25% of the retrieved studies with the blind on. Both reviewers (the author, and the second reviewer) conducted the first screen independently on Rayyan using its three options "Included", "Excluded", and "Maybe". The Cohen's Kappa was calculated to show the level of agreement (reported in section four), after which a follow up meeting was arranged to discuss any potential conflicts, and to discuss which studies should be included for the second screening of full texts.

Following the review meeting, as Newman and Gough (2020) propose, the author conducted a second screen of the full texts labelled as "Included", or "Maybe" to ensure the selected studies met the inclusion criteria. Quality assessment of the studies, which is often carried out in systematic reviews, was not conducted for two reasons: this study is not a systematic review, and the study is, in itself, a quality assessment of the published applied linguistics review.

Once the studies had been second-screened, the final selection of the studies that met the inclusion criteria was subsequently coded, or mapped, in a systematic map to ascertain how research in the studies had been conducted (Newman & Gough, 2020). Macaro (2020) proposes multiple items that should be included in a systematic map, however, since Newman and Gough (2020) highlight that purpose of the coding is to "allow assessment of the quality, and

relevance of the studies in addressing the review question" (p. 12), only the data that could be informative in terms of answering the RQs of this paper was coded. This was captured in a data extraction table that was completed for each study separately, an example of which is Table 2.

Study number	
Authors and year of publication	
Title	
Research question(s)	
Participants	
Setting	
Methodology	
Gender as a variable	
Method of collecting information on gender	
Results (gender)	

Table 2: Data extraction table used for systematic mapping of information

3.3 Corpus analysis – RQ1

The collated studies represented a corpus which meets the three criteria described by Rose et al. (2019): "the texts included are (1) authentic (i.e. naturally occurring), (2) representative (i.e. sampled in a principled manner from a larger population of texts) and (3) machine-readable (i.e. can be analysed using software)" (p. 218). The specialised corpus was "compiled systematically and for a pre-specified purpose" (Rose et al., 2019, p. 218). Biber (1993) highlights the need for a corpus to be representative of the wider population which was achieved by the systematic approach to its development, thus

ensuring the necessary rationale, and the description of the process are provided (Nation, 2016). The systematic approach also defined the size of the corpus.

A text in the corpus is defined as an individual study. Although Sinclair (2004a) recommends the use of complete texts, Coxhead (2020) notes that refences in journal articles would lead to little gains in terms of their analysis, which is why they were removed in the preparation, and cleaning process. Preparing, and cleaning corpora ahead of their analysis is important (Coxhead, 2020; Nation, 2016), as it ensures no typographical errors are present, and that the corpus analysis software can process the corpus without any problems. The preparation process was informed by Sinclair (2004b), and it was carried out in the following steps:

1) Each text was saved as a PDF which had been retrieved through the Bodleian library.
2) References were deleted from each PDF as whole pages, although notes, and appendices were kept in case they contained the definition of gender. Where the references started on the same page as the end of the study's conclusion, these pages were retained. This was done manually using the Preview app available on Apple computers.
3) The text files were relabelled to ensure clear traceability of findings during the analysis, as well as to avoid bias by anonymising the studies. These labels were recorded in the systematic map.
4) The PDF files were converted into a plain text format (.txt) using AntFileConverter (Anthony, 2022).
5) The plain text files were manually scanned for unusual characters, or unnecessary spacing in the word "gender" to ensure the corpus analysis software could reliably read them.
6) A back-up of the processed corpus was saved on the cloud in case the working file got corrupted.

To answer RQ1, a concordance analysis was used using the term “gender” as the search term to review its pattern of use to ascertain the presence of the definition of gender (Rose et al., 2019). This corpora analysis tool had been selected to answer RQ1, since the definition of gender could have been present anywhere in the studies, not only in their methodology sections. The Key Word in Context (KWIC) function of AntConc (Anthony, 2023) was used for this purpose, and the surrounding text was qualitatively evaluated in the concordance output (Rose et al., 2019) for what could resemble a definition of gender. The results were sorted by frequency, since it was presumed a definition could be found focusing on collocations, such as “gender is”, “gender can be”, or “definition of gender”. However, to ensure no alternative phrasings were used, every result was reviewed regardless of context. If the presence of a definition was indicated by the context in KWIC, the File View function was used to read the complete sentences containing the concordance output to confirm this. The section of the document the definition of gender was present (if applicable), was also obtained through the File View function of the corpus analysis software used, AntConc (Anthony, 2023).

3.4 Qualitative text analysis – RQ2

To answer RQ2, the methodology sections of each included study were read by the author to ascertain whether they contained information on how the information on participants’ gender was collected. This was carried out at the same time as the systematic mapping, and the data was extracted in the systematic map itself, before being pooled for presentation in the results section. Studies were grouped according to how gender was measured, and frequency counts of these grouping alongside exemplar excerpts provided insights into its measurement.

3.5 Qualitative text analysis – RQ3

3.5.1 Selecting studies for RQ3

This study concerns itself with analysing how *gender* is treated within the field of applied linguistics. It hypothesises that the research within the field does not do due diligence to differentiate the concept from the concept of biological *sex*, which can be one of the contributing factors leading to inconclusive outcomes of the research field. That is why the studies which claim that gender as a variable does have an effect on the outcomes of their research were of high interest to this study, and why statistically significant findings were used as the selection criterion for narrowing down the studies for RQ3. This approach of conducing the qualitative text analysis on a narrower selection of studies had been selected to ensure this analysis could have been given sufficient attention (Macaro, 2020).

3.5.2 Inductive approach

Rose et al. (2019) outline two approaches to qualitative content analysis: a qualitative text analysis, and a thematic text analysis. The former uses a pre-established coding framework pre-established around themes central to the research which should be developed, and piloted before being used in a study. In thematic text analysis, on the other hand, the themes emerge. This study did not use a pre-established coding framework, and instead, it let the themes emerge. The purpose of the qualitative text analysis was to ascertain the extent to which the results of individual studies could be interpreted in line with the definition of gender, as it is important to consider whether the findings, and their discussion match the methodology used (Macaro, 2020). Methodology, in this case, pertains to the presence of a definition of gender.

To achieve this, the entire studies selected for this level of analysis were read, and key themes in their analysis were extracted, and studies were grouped according to similar themes in their treatment of gender. In the discussion, the

findings were used to identify any potential gaps in research, and its methodology (Macaro, 2020).

3.5.3 Definition of gender

For the purposes of this study, a working definition of gender needed to be established. This paper sees two principal ideas as key to the definition. The first is drawn from Butler's (2006) idea of gender being a behaviour, i.e., something that one does. This idea is also cited in the post-modernist definition of the term *gender* in the *Longman dictionary of language teaching and applied linguistics* which posits that "gender is viewed more as a process (something that someone does or performs in interaction, rather than an attribute that one possesses" (Richards, 2013, p. 240). The second idea, as discussed in the literature review, is that gender is culturally conditioned, in that it is a result of socio-cultural expectations, which stands in contrast to the term *sex* which pertains to biological differences between men, and women. This working definition of gender was used to interpret the findings of the studies selected for this level of analysis. Where the included studies had been found to include a definition of gender (subject to results to RQ1), a further layer of analysis was conducted to ascertain whether their results were discussed in line with that definition.

3.5.4 The context of studies

The context of the individual studies examined in this study was also considered. Firstly, Rose et al. (2019) note that in qualitative text analysis, documents should be interpreted within the contexts in which they have been created to ensure that the author's intentions are interpreted correctly. Data collection was the sole method of this study because since the documents used were academic journal articles, as discussed in the literature review, the authors should do their due diligence in their methodology, and should be precise in their use of language. Secondly, the context of the studies is highly pertinent to RQ3, and the discussion of the findings of this study, because gender is culture-

dependent, and culture is context-dependent. In other words, culture is a part of context, and thus context can have an impact on how gender is being interpreted by the researchers in their studies. This made context the third element of the definition of gender.

3.6 Ethical considerations

Coxhead (2020) notes that researchers need to consider ethics in corpus construction, with Rose et al. (2019) highlighting that consent should be obtained before using the data. Consent had not been obtained prior to commencing this study, since the study is a systematic content analysis of published research which is publicly available through Scopus. As such, the inclusion of the studies in this review does not differ from using them as references in any other academic study. For copyright reasons, the corpus that was created in this study is not publicly available, and the access to it is limited to the author, and the dissertation supervisor. The corpus was used only to analyse the selected studies, and it does not infringe on the intellectual property rights of the authors of the individual studies contained within it. However, a list of the included studies is included in the appendix (Appendix 1) of the dissertation to bring transparency to the contents of the corpus, and will allow external researchers to compile a similar corpus with relative ease.

4. Results

4.1 Included studies

The search was conducted on the Scopus database on 31 January 2024, and it returned 98 studies which were exported to Rayyan for screening. One duplicate was identified, bringing the total of studies for title/ abstract/ keyword screening to 97 studies. Figure 2 provides an overview of the study selection process based on a PRISMA 2020 flow diagram (Page et al., 2021). Although this is not a systematic review, a PRISMA flowchart is included to provide a fully transparent, detailed overview of how the studies for this research had been selected (Petticrew & Roberts, 2006), not only so that this study could be replicated, but also to enhance its methodological credentials.

The first screen conducted by the author resulted in 55 exclusions, 25 inclusions, and categorised 17 studies as "Maybe". The second reviewer screened a random selection of 25 studies with the blind on, and they excluded 14 studies, included five, and labelled six as "Maybe". A Cohen's Kappa was calculated ($k = (p_0 – p_e) / (1 – p_e)$), and its value of $k = 0.8376$ indicated near perfect agreement. A post-screen meeting was conducted on 29 February 2024, in which five conflicts were resolved. An over-inclusive approach was agreed on, meaning where the reviewers did not have enough information to decide whether to include, or exclude a study, the study in question was selected for the second, full text-screening. All studies labelled as "Maybe" were subject to the second screen.

42 studies were thus selected for the second screen of full texts. Full texts were retrieved either from the Bodleian Library's SOLO, directly from the journal websites, or via interlibrary requests. No critical appraisal of studies was conducted at this stage, since the purpose of the research itself was to critically examine the methodological rigour of existing research, and to do so in a methodical way. Resulting from the second screen, 36 studies met the inclusion

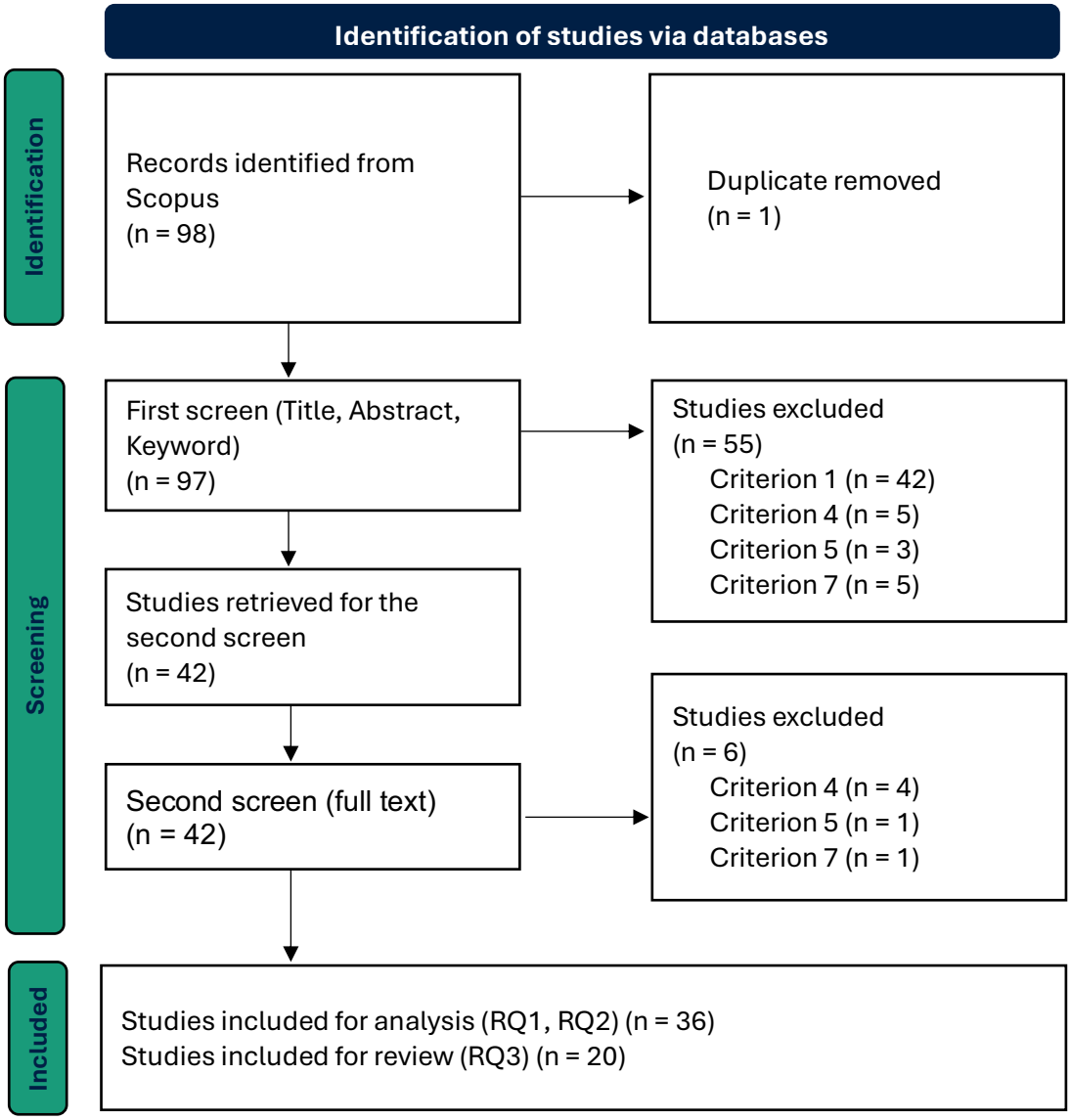

Figure 2: PRISMA flowchart for study selection

criteria, out of which 20 found a statistically significant, or—in the case of qualitative studies—a notable effect of gender on the dependent variable, meaning they were relevant for RQ3. The studies were sorted in an alphabetical order of the authors' surnames, and subsequently renamed both to enable easier working with the files in the corpus software, and in this research, as well as to reduce research bias by anonymising the studies. Hence, in alphabetical order, study one became S1, study two became S2, and so forth.

The full bibliographical information of the 36 studies can be found in Appendix 1 – as Petticrew and Roberts (2006) highlight the importance of managing references in a research of this type, reference management software Mendeley was used to this effect.

Once the screening was finished, and the studies renamed, data extraction was conducted. Appendix 2 includes the data extraction table containing the information on the basic characteristics of the included studies, their research questions, and basic methodological approach/ instruments used, as well as the data relevant for RQ2. This data extraction also served to identify the presence of significant/ notable findings pertaining to gender, which informed the selection of studies for a narrower analysis for RQ3.

4.2 Characteristics of the included studies

Focusing on the characteristic of the included studies, as can be seen in Figure 3, most of the included studies had been conducted in Asian countries (n = 25), with 12 of them being from Iran. Seven studies had been conducted in Europe, four in Northern America, one in Australia, and one study did not specify the country in which their research had been carried out. There were no studies from Africa, nor from Southern America. Most studies, as can be seen in Figure 4, had been conducted at universities (n = 18), nine at language schools, which were sometimes referred to as “language institutes” in the studies, and four studies used a pool of volunteers drawn from the general public. One study was conducted at both universities, and language schools, and another one at language, and state schools. “Other” setting includes one study conducted at “educational centres”, and one at “community colleges”. One study did not specify its setting. In total, 38,836 participants participated in the 36 studies. One of these studies included 29,883 participants, with the other 35 contributing

8,953 participants. As can be seen in Figure 5, the focus of the studies varied greatly, and the body of research covered a wider range of topics.

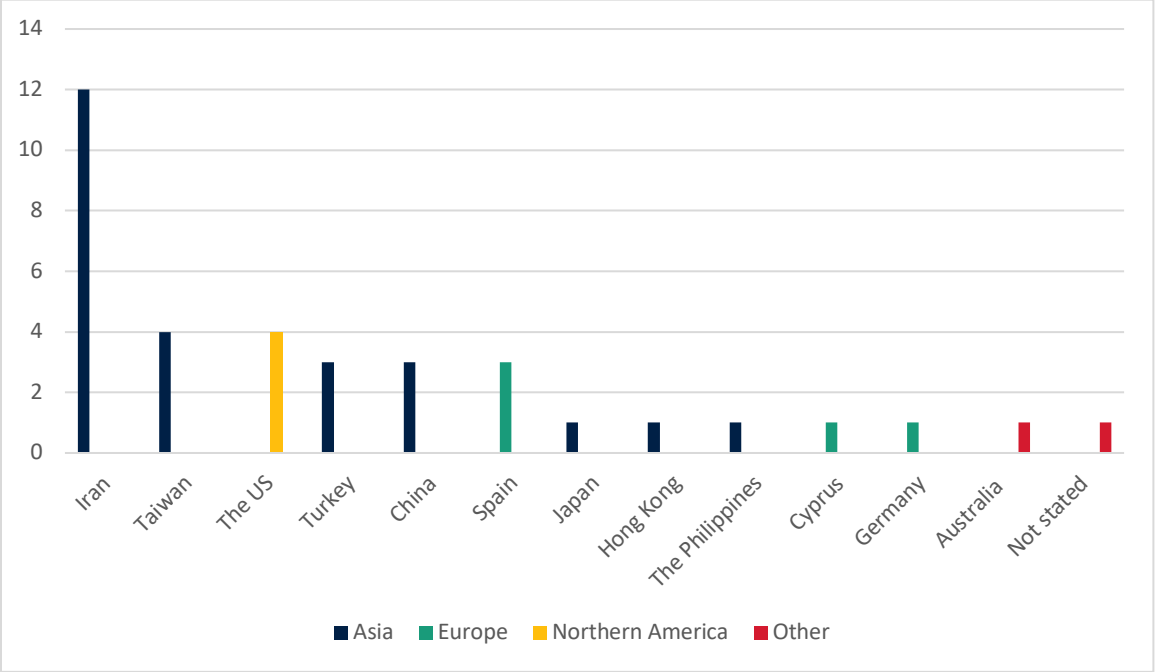

Figure 3: The location of the included research

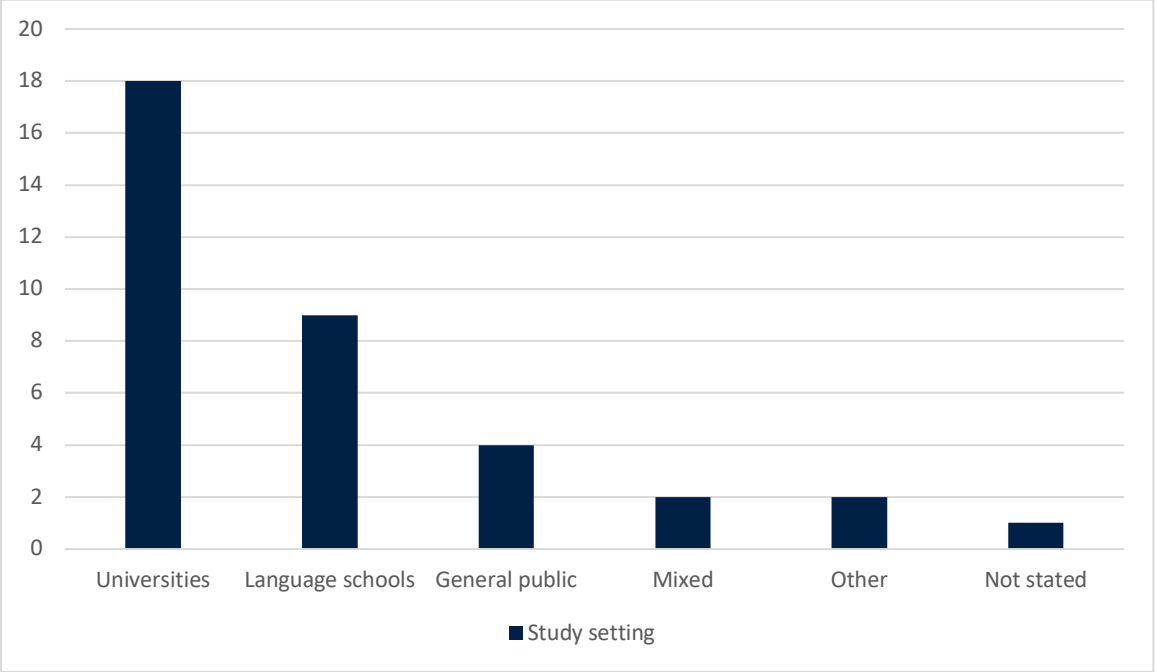

Figure 4: The setting of the included research

Study	Focus
S1	Willingness to communicate
S2	Uptake of Language Massive Open Online Courses (LMOOC)
S3	Perceptions of Kahoot
S4	Acquisition of L2 vernacular variation
S5	Interlanguage pragmatic learning strategies
S6	Vocabulary retention/ vocabulary learning strategies
S7	Second language anxiety
S8	Teacher hopelessness
S9	Gender identity shifts
S10	Teacher burnout
S11	Foreign language anxiety
S12	Language learning strategies
S13	Pragmatic competence
S14	Compliment response patterns
S15	Gender stereotypes, and language attainment
S16	Attention, hyperactivity, and impulsivity
S17	Learning styles
S18	Disfluency types
S19	Out-of-school contact with English
S20	Perceptual learning styles, and attitudes towards communicative language learning
S21	Listening comprehension
S22	Item detection in the Peabody Picture Vocabulary Test
S23	Positive feedback rate, and course level promotion
S24	Teacher education philosophy, and teachers' professional knowledge
S25	Cognitive, and metacognitive strategy use
S26	Post-teaching activity types, and vocabulary learning

S27	Hypertext learning experience
S28	First impressions of language teachers
S29	Emotional intelligence, and vocabulary strategy use
S30	Willingness to communicate
S31	Academic performance
S32	Language learning strategies
S33	Attitudes towards smartphone-based learning
S34	Classroom activity preferences
S35	Biliteracy, and cognitive skills
S36	Sources of self-efficacy

Figure 5: The focus of the included research

4.3 RQ1: Presence of a definition

To answer RQ1(Does current research examining gender in applied linguistics include its definition?), a corpus analysis using AntConc (Anthony, 2023) was conducted. The 36 included studies were converted into plain text files using AntFileConverter (Anthony, 2022) before being used to create a corpus. A manual scan of the converted plain text files was carried out to identify any anomalies that could have potentially occurred during the file conversion. S2 was found to have included a significant number of strange characters, hence an additional, manual search for the term “gender” using command + f in a PDF reader was conducted for this study only. This is reported here to ensure methodological rigour and transparency, as Petticrew and Roberts (2006) note that methodological changes need to be documented.

The corpus of the 36 studies included 253,097 tokens. The term gender had 820 hits. Using the KWIC tool, the corpus was ordered by frequency, sorted to left, and set to display 10 tokens on either side of the hit. The hits were manually reviewed, scanning the immediate context for a phrase that could indicate the presence of a definition of gender. Subsequently, the corpus was sorted to right,

and the scan was repeated. The concordance output is illustrated in Figure 6 – there were 100 hits on each page, so a total of nine pages (820 hits) were scrolled through twice (sorted to left, and right).

These scans indicated that only one study (S5) included an explicit definition of gender, which can be seen in Figure 6. This was present in the discussion section of the paper, where the authors were interpreting their results, and they stated that "for one thing, gender should be viewed from the angle of its broader sociocultural definition, i.e., the gender-as-a-social-pattern phenomenon (Kasper & Rose 2002), not from the one-dimensional biological perspective" (Derakhshan et al., 2023, p. 18).

S9 also problematised what *gender* meant, acknowledging "that gender is constructed along with other identity categories such as class, race, and linguistic and cultural background (Eckert & McConnell-Ginet, 1992)" (Gordon, 2004, pp. 439-440). This was done to justify the use of, and define the meaning of their dependent variable (gender identity), however, gender itself remained undefined in this study.

For added rigour, the corpus was also ordered by value instead of frequency, and the scan repeated with the results sorted both to left, and right. Subsequently, the context size was increased to 20 tokens, and reviewed again as ordered by both frequency, and value, sorted left, and right, resulting in eight manual reviews of the corpus in total. The corpus was additionally searched for the terms "define", "defined", and "definition". None of these additional searches indicated the presence of the definition of gender.

Therefore, the answer to RQ1 is that only one of the included studies which claim to be researching gender in applied linguistics included a definition of gender. This supports hypothesis one in that most current research in applied linguistics investigating the concept and/ or impact of gender does not include a definition of the concept.

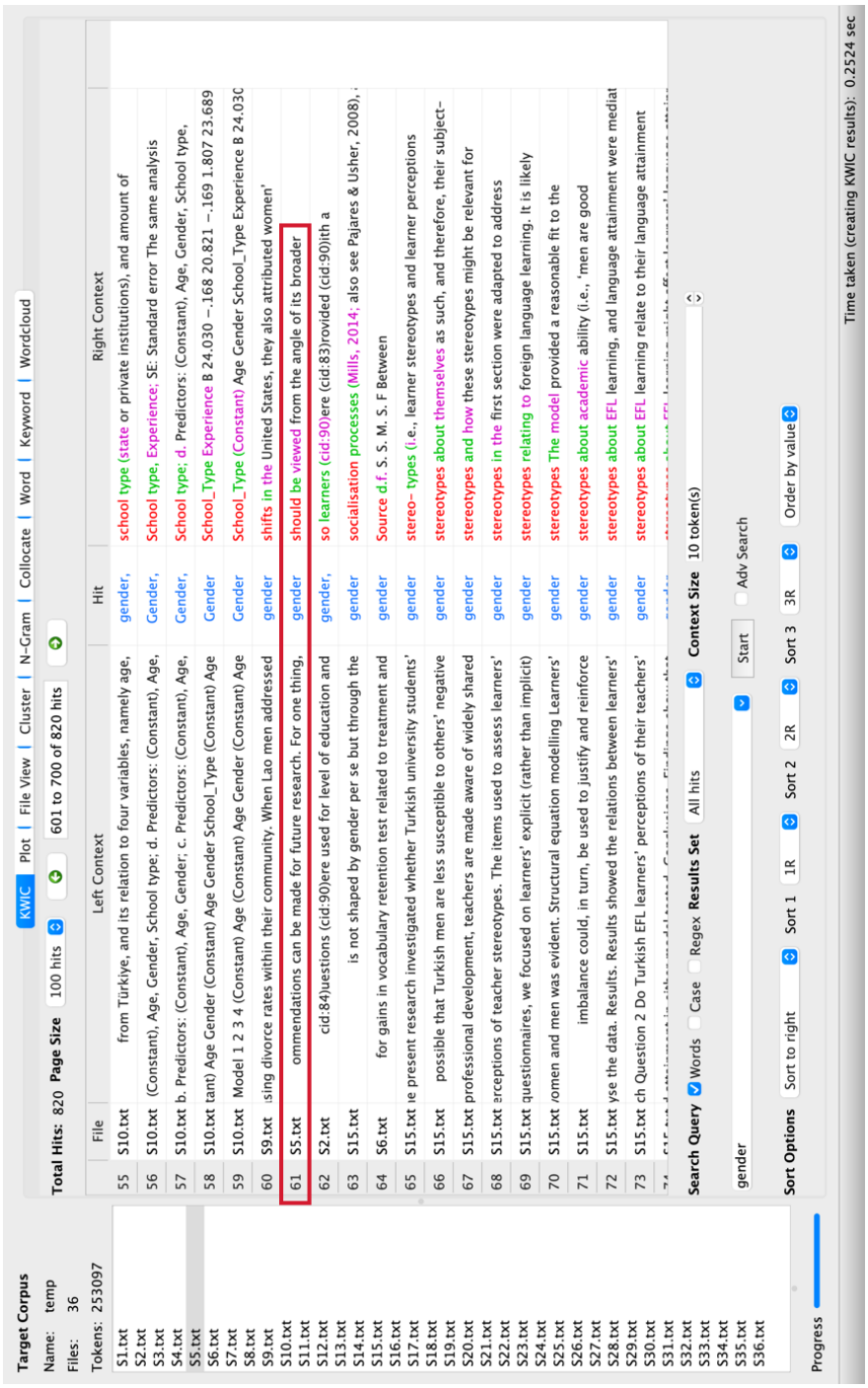

Figure 6: Concordance output

4.3 RQ2: Presence of a valid method

To answer RQ2 (Does current research examining gender in applied linguistics include a valid method in which the information on the participants' gender was collected?), the method in which the authors of included studies collected the information on participant gender was manually identified, and extracted into the data extraction table (see Appendix 2 for more detail). As illustrated in Figure 7, 12 studies used a questionnaire to collect the information on participants' gender. One of these studies (S2) offered their participants three multiple-choice options ("female", "male", and "other"), and three studies (S4, S11, S36) measured gender on a dichotomous ("female", "male") scale. The other eight studies which used a questionnaire to ask about the participants' gender (S1, S5, S7, S8, S12, S15, S27, S33) did neither provide the questionnaire in the appendix for this information to be ascertained, nor did they provide information on how the gender question was asked. However, it could be speculated that gender was measured on a dichotomous scale since the participants were referred to as "male" or "female" in all 36 studies. Similarly, out of the total of 36 studies, all but seven (S2, S6, S12, S20, S21, S23, S24) stated the participant gender composition on a dichotomous scale, with the said seven not stating the gender mix at all.

A further four studies (S10, S19, S31, S34) indicated that they required demographic/ biographical information, where gender could have been included. However, only one of these three studies (S31) specified that the demographic information included gender, with the others providing no detail on what sort of demographic/ biographical information had been collected. In the three other cases (S10, S19, S34), positing that the collected information included information on gender is a mere speculation, and it cannot be viewed as collecting information on the participants' gender using a valid method.

19 studies (S3, S6, S9, S13, S16, S17, S18, S20, S21, S22, S23, S24, S25, S26, S28, S29, S30, S32, S35) did not state how the information on participants' gender was collected. Four studies detail that the participants had been selected based on their gender (S14, S18, S21, S29), often to include an even mix of females and males in the studies. As the information on participants' gender had not been captured, however, the methods used to select the participants remained unexplained. An exception to this was study S14 which stated that participants "were required to mention their gender and age in the specified blanks in the questionnaire" (Khaneshan & Bonyadi, 2016, p. 762) (open choice). However, this assertion needs to be questioned, since the questionnaire provided in the appendix of the study provides participants with a dichotomous choice, not a blank to fill in.

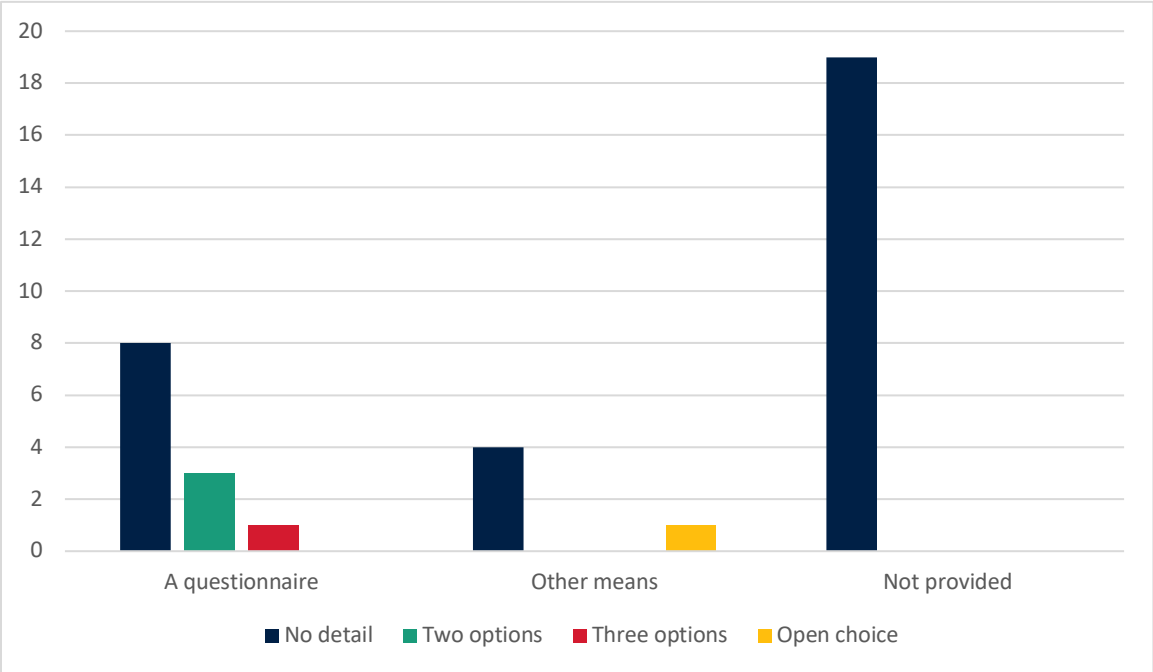

Figure 7: The method of collecting information on participants' gender

A further analysis was conducted to ascertain whether there was any relationship between the context of the studies, and the methodology used to collect the information on participants' gender. This was done by manually populating a table with the data obtained for RQ2 (see Appendix 3). A noteworthy

finding was that all four studies which chose participants based on their gender (S14, S18, S21, S29) were conducted in Iran, which will be discussed in more detail in section five of this dissertation.

In summary, out of 36 examined studies, 14 required information on gender from their participants, 19 studies did not state how this information was collected, and three studies collected biographical information which may have included information on participants' gender. Out of the 14 studies which asked for gender information, five studies clearly stated how the information was collected: one study provided participants with three options to choose from, three studies gave their participants two options, and one study claimed to be using an open-ended question. The other studies did not report the method of collecting this information in sufficient detail, and thus could not be evaluated. The breakdown of the study numbers is provided in Table 3.

Studies which asked for information on participants' gender (n = 14)	
A questionnaire: three options (n = 1)	S2
A questionnaire: two options (n = 3)	S4, S11, S36
An open choice (n = 1)	S14
A questionnaire: no detail provided (n = 9)	S1, S5, S7, S8, S12, S15, S27, S31, S33
Studies which did not ask for information on participants' gender (n = 19)	S3, S6, S9, S13, S16, S17, S18, S20, S21, S22, S23, S24, S25, S26, S28, S29, S30, S32, S35
Studies which asked for bibliographical information (n = 3)	S10, S19, S34

Table 3: Summary of studies for RQ2

Therefore, the answer to RQ2 is that five of the included studies which claim to be researching gender in applied linguistics stated using a valid method of collecting information on participant's gender, although what constitutes a valid method will be further discussed in section five of this dissertation. This finding supports hypothesis two in that most current research in applied linguistics investigating gender does not stipulate how the information on participants' gender was captured, and in some cases, participants had been selected based on their perceived gender identity by the researchers.

4.4 RQ3: Interpretation

4.4.1 Included studies and approach

To answer RQ3 (To what extent can the findings of current research examining gender in applied linguistics be interpreted in line with the definition of gender?), only the studies which report a significant, or a notable effect of gender on their dependent variable had been selected for an in-depth critical analysis. From the second screen, 20 studies in total reported either a statistically significant effect of gender, or, in the case of qualitative studies, a notable effect of gender (see Appendix 2 for the results of these studies). The studies selected for this level of analysis were: S2, S4, S6, S9, S11, S14, S15, S16, S17, S18, S19, S22, S26, S27, S28, S29, S31, S32, S33, and S36. These studies were read in full, and their findings were interpreted according to the extent to which they were in line with the definition of gender posited in this study, or not. As outlined in the methodology section, this definition included three elements, namely that (1) gender is performed in an interaction, (2) it is subject to socio-cultural expectations, and (3) these can be context-dependent. The methodology of the studies was considered in this interpretation, too, as methodology can have an impact on research findings. It needs to be further noted that this dissertation focused on interpreting the results of the included studies in line with the definition of gender only. If a study was found not to be in line with the definition of gender, their findings could have been subject to the influence of a different

variable, such as sex, cognitive abilities, or other factors. Whilst these potential variables were considered in the analysis, it was beyond the scope of this thesis to attempt interpreting the findings in line with the definitions of those concepts, or attributing the findings to other variables.

Following an inductive approach to qualitative text analysis, some key themes emerged, and the following section is organised in the following way: first, studies whose findings are in line with the definition of gender are mentioned, followed by studies whose findings are not in line with the definition of gender. The impact of methodology on the results is then considered, before focusing on studies researching LLS use, and learning styles, where the analysis was more problematic. Each section includes a more in-depth critique of one study to either illustrate how the findings were interpreted, or to problematise the findings/ their interpretation. An overview of the qualitative text interpretation can be found in Appendix 4.

4.4.2 In line with the definition of gender

The following 10 studies were found to be concordant with the definition of gender: S2, S4, S9, S11, S14, S15, S18, S19, S28, S33. The studies examined topics, such as uptake of courses (S2), acquisition of L2 vernacular variation (S4), gender identity shifts (S9), foreign language anxiety (S11), compliment response (S14), perceptions of gender stereotypes (S15), dysfluency types (S18), out-of-school contact with English (S19), first impressions of language teachers (S28), and attitudes towards smartphone-based reading (S33). Although the design of the studies did not always include an interaction (S2, S11, S14, S15, S19, S28, S33), which is pertinent to the definition of gender used in this study, they investigated something that happens in, or is borne out of an interaction. Furthermore, the influence of socio-cultural expectations, and/ or the role of context were considered important in having an effect on the findings of these studies.

To illustrate, study S14 investigated how participants of different gender, and age used compliment response strategies differently (Khaneshan & Bonyadi, 2016). Compliment responses are an aspect of a speech act – they are conversational, potentially formulaic responses to compliments (Pomerantz, 1978, as cited in Khaneshan & Bonyadi, 2016). To investigate what strategies participants used to express compliment response, 100 advanced Iranian EFL learners were selected based on their Cambridge First Certificate in English (FCE) reading, and writing scores. The participants subsequently completed a discourse compliment task in which they provided responses to compliments in hypothetical scenarios. These responses were then analysed using a framework of compliment response strategies. The mixed-method analysis of the responses indicated that, pertinently to gender, there were no noticeable differences between the different strategies used by males and females, and that male and female participants differed mostly in terms of the words they chose to enact these strategies. These findings could be interpreted as concordant with the definition of gender because, as the authors themselves state, "compliment responses can be the demonstration of the social-cultural values and politeness varieties of the speakers" (Khaneshan & Bonyadi, 2016, p. 761) which are context-dependent, since they depend on the socio-cultural norms of a given context. Compliment response strategy use is also dependent on the characteristics of the other speaker (their age, seniority, gender), since people respond to compliments in interactions.

Other studies were interpreted in a similar way. As illustrated in Table 4, which is extracted from Appendix 4, study S11 was found to be concordant with the definition of gender, since foreign language anxiety is something that occurs in an interaction, and socio-cultural expectations with regards to such interactions can not only exist, but they are likely to be context-dependent.

Study	S11
Context	Japan, university
Focus	Foreign language anxiety
Methodology	A questionnaire (self-assessment)
Significant findings	Females had higher levels of anxiety than males (t (250) = -2.893, *p* = 0.004) Females spent more years studying English than males (t (250) = -2.288, *p* = 0.023)
In line with the definition of gender	✅ Anxiety in foreign language use/ classroom is something that one does in an interaction, and hence it could be socio-culturally influenced. This can be context-specific. *The method does not include an interaction between participants.

Table 4: Interpretation of study S11

4.4.3 Not in line with the definition of gender

The following seven studies were found not to be in line with the definition of gender: S6, S16, S17, S26, S27, S29, S31. The studies investigated the following topics: the effect of vocabulary learning strategies on vocabulary retention (S6), the effect of ADHD on performance in a proficiency test (S16), learning style preferences (S17), the effect of post-teaching activity types on vocabulary learning (S26), perceptions of hypertext learning experience (S27), vocabulary strategy use (S29), and the impact of pre-admission English scores on academic outcomes (S31). The majority of these topics do not involve an interaction between language users, and the study design, in most cases, did not involve interaction between participants. Some interaction between participants was likely to have occurred in studies S6, S26, and S31, as the first two included an intervention where learners would have worked together, and S31 was a longitudinal study where the participants would have naturally interacted with

each other over the course of the study, thus being unable to isolate it. However, the results of these seven studies were viewed as not being in line with the definition of gender, because socio-cultural expectations are unlikely to have an effect on the variables they are investigating.

To provide an insight into this analysis, study S16 investigated the impact of attention deficit, hyperactivity, and impulsivity symptoms on EFL test performance, and the gender differences within that (Liang & Kelsen, 2017). 229 EFL learners at a Taiwanese university completed an 18-item, adult ADHD Self-Report Scale (ASRS), the scores from which were compared against the scores from the Soochow University English Proficiency Test, which was administered in weeks 15, and 16 of the spring term. The results, which were sorted by how likely the participants had ADHD, revealed that one item on the ASRS was statistically significant for gender ($p = 0.035$). With regards to their EFL test performance, likely ADHD females performed the best, and likely ADHD males scored the lowest, with both females, and males who were unlikely to have ADHD scoring in the middle. What is more relevant to this dissertation is that ADHD is a psychiatric diagnosis, and although it can affect a person's behaviour, it is not performed in an interaction, nor is it influenced by socio-cultural expectations. As the authors state, ADHD can be treated (Liang & Kelsen, 2017), which means it is more likely to be a biological factor, and hence cannot be interpreted in line with the definition of gender.

Similarly, study S27 was not found to be concordant with the definition of gender because reading is not performed in an interaction, and this internal process is unlikely to be influenced by socio-cultural expectations. This analysis process is outlined in Table 5, which is extracted from Appendix 3.

Study	S27
Context	Taiwan, university
Focus	Hypertext learning experience
Methodology	Hypertext reading task + a perceptions questionnaire
Significant findings	Females felt that they were able to reduce reading time in hypertext use more than males (p = 0.044), and found it easier to do the task than males (p = 0.019) Females scored higher on hypertext learning experience than males (d = 0.64)
In line with the definition of gender	❌ Reading is not done in an interaction, and despite the existence of gender stereotypes, reading is not a socio-culturally conditioned behaviour. This could be context-dependent, with context having an impact on the availability of reading material, but this contextual factor would not be related to gender.

Table 5: Interpretation of study S27

Three studies have not been referred to yet (S22, S32, S36), as they were found to be partly in line with the definition of gender, and partly not. These will be addressed in the following two sections.

4.4.4 Impact of methodology

Methodology – the way the investigated studies had been designed – was found to have a notable impact on the findings of the studies, and to what extent they could be interpreted in line with the definition of gender.

12 studies used questionnaires, or other, self-reported scales (S2, S11, S15, S16, S17, S19, S27, S28, S29, S32, S33, S36). As such, self-reported methods, it can be argued, are to some extent susceptible to socio-cultural expectations for participants to respond either the way they would like to be perceived, or the

way they believe they are expected by the researchers to respond. This can be highly context-dependent, too. For instance, study S19 investigated how English language learners use English outside of lessons, which could not only vary by context (i.e., what might be popular in any given country, or a region), but the study also used a questionnaire to collect the data, which might have meant that the participants provided the same answer as their peers to conform to socio-cultural expectations.

Longitudinal studies, and studies employing a delayed post-test also need special attention. Whilst the findings of study S9 were found to be in line with the definition of gender, this was because the study utilised a range of qualitative data collection methods which were focused on observing behaviour, and interactions in a particular context. The same could not be said for a longitudinal study S31 which investigated the possible correlation between pre-admission English scores, and cumulative grade point average (GPA) on a four-year pharmacy programme (Tenney et al., 2020). The quantitative results from 113 participants indicated that the correlation was only significant for females (p = 0.009). This was interpreted as not being in line with the definition of gender, because tests results are not performed in an interaction, and socio-cultural expectations have no direct impact on what answers participants supply in content tests. Having said that, socio-cultural expectations might have had an impact on the participants' behaviour with regards to, for example, how much effort they put into their learning during the course, and such expectations could be context specific: people of a particular gender in one culture might feel more pressure to study hard than in others. Without further qualitative data to provide an insight into the potential impact of these expectations, which this study lacked, it is challenging to isolate cultural influences, from the cognitive ones, hence it cannot be argued that the results are in line with the definition of gender. Similarly, as can be seen in Table 6, studies S6, and S26 utilised delayed post-tests after their interventions, meaning that the performance in the tests,

which would not be interpreted in line with the definition of gender by itself, could have been affected by socio-cultural expectations. These findings pertaining to longitudinal designs/ study designs using delayed post-tests without qualitative elements hint at the importance of using the right methodology to be investigating the concept of gender.

Study	S6	S26
Context	Iran, language school	Iran, language school
Focus	Vocabulary retention/ vocabulary learning strategies	Post-teaching activity types, and vocabulary learning
Methodology	Experimental (pre-test, intervention, post-test)	Experimental (pre-test, intervention, post-test)
Significant findings	Females in the experimental group improved their vocabulary retention (p = 0.05)	Gender had a significant impact on vocabulary learning: female participants outperformed males (p = 0.000)
In line with the definition of gender	✗ This is not performed in an interaction, the strategies are not used under socio-cultural influence.	✗ Vocabulary acquisition is more cognitive than behavioural. Whilst some activity types included interaction, and the context (single sex experimental groups) could have had an impact on the learners' behaviour because of socio-cultural expectations, the post-test was delayed, so the

		interaction in itself cannot be isolated as a correlative factor.

Table 6: Interpretation of studies S6, and S26

The importance of using the right methodology to examine gender as a variable was further noted in studies S22, and S36, which can be interpreted in line with the definition only in part. Study S22 investigated the performance of females, and males on the Peabody Picture Vocabulary Test IV (PPVT). The study found that 3% of items showed a difference in detection by gender, and that males scored significantly higher than females ($t = 5.4$; $p < 0.001$) (Pichette et al., 2019). As such, the PPVT is designed to be conducted individually with the researcher, and stopped at the point of saturation, i.e., when the participant cannot answer any more questions because of the increased difficulty of the test items (Pichette et al., 2019). Under such circumstances, the test does not include an interaction, and it would be neither subject to socio-cultural expectations, nor context-dependent; it could not be said to reflect the definition of gender. In this study, however, the researchers conducted the test in groups, and continued until the end of the test, thus allowing learners to guess the answers to the items they did not know. Continuing beyond the point of saturation, and allowing the participants to guess the answers was a purposeful study design decision to allow males, who have been found to be more willing to take risks, to guess the answers (Pichette et al., 2019). However, as the authors acknowledge, risk-taking is socio-culturally influenced because of male propensity for this behaviour. The implication this had on the study design was that the data obtained up to the individuals' point of saturation was not concordant with the definition of gender, and it might have been more about cognitive abilities/ language knowledge, whilst the data obtained whilst participants could guess the answers did fulfil the socio-cultural aspect of the definition of gender. As a result, the findings could be partly interpreted in line with the definition of gender.

A part of the findings of study S36 could also be interpreted in line with the definition of gender. The study investigated the variance caused by four different sources of self-efficacy on English public speaking self-efficacy for males, and females. The four sources were enactive mastery experience (i.e., one's own perception of their abilities), vicarious experience (i.e., comparing one's performance with others), verbal persuasion (i.e., feedback from others), and physiological and affective states (i.e., one's ability to manage their emotional, and physical stress reactions) (Zhang & Ardasheva, 2019). Significant variance between genders was reported for the first three sources. However, only vicarious experience, and verbal persuasion could be interpreted in line with the definition of gender because comparing oneself to others, and getting feedback from others are performed in interactions, and as such they could be subject to socio-cultural expectations according to which this is done. The way people compare themselves, and how they deliver feedback is also context-dependent, as some cultures/ contexts might be more direct, others more reserved, meaning all three elements of the definition of gender are met for these two sources of self-efficacy. Enactive mastery experience, on the other hand, is a self-evaluation ability, meaning it is not performed in an interaction, and context, and socio-cultural expectations are unlikely to have significant influence on it. Since the three elements of the definition of gender are not applicable to this source of self-efficacy, it could not be interpreted as concordant with the definition of gender. Therefore, two thirds of the findings of this study could be interpreted in line with the definition of gender, and one third could not.

4.4.5 Language learning strategy use, and learning styles

Four studies investigating LLS use (S6, S29, S32), and learning styles (S17) deserve special mention. Studies S6, and S29 were found not to be in line with the definition of gender, since LLSs investigated in those studies were not performed in interaction, and thus were unlikely to have been influenced by

socio-cultural expectations. However, unlike studies S6 and S29, study S32, which investigated gender differences in EFL learners' LLS use, did provide a breakdown of its results by six LLS categories, owing to its use of Oxford's (1990) Strategy Inventory for Language Learning (SILL) (Tercanlioglu, 2004). This detail is pertinent, because some strategy categories, such as "Managing your emotions", or "Learning with others", are performed in interaction, and their use could be affected by socio-cultural expectations. Because of this, it could be said that the strategies investigated by study S32 could be partly interpreted in line with the definition of gender. The other strategy categories ("Remembering more effectively", "Using all your mental processes", "Compensating for missing knowledge", and "Organising and evaluating your learning") were not found to be concordant with the definition of gender, and they could be tapping into cognitive/ biological aspects of a person instead.

Study S17, which investigated learning styles, was interpreted as not being in line with the definition of gender, because learning styles are not performed in an interaction, and are unlikely to be subject to socio-cultural influences. However, the study did report notable differences in learning styles between Hispanic males, and Asian males: Hispanic males stated a preference for the read-write ($p < 0.001$), or kinaesthetic ($p < 0.05$) styles, whilst the latter group were more aural, or read-write. These cross-cultural differences suggest that context might play a role in determining learning styles, however, on their own they might not be enough to interpret the findings in line with the definition of gender. This is because manifesting a particular learning style is not performed in interaction which is key to the definition of gender, and learning style preferences could be more cognitively/ biologically driven, rather than being influenced by socio-cultural expectations.

4.4.6 Summary

Therefore, the answer to RQ3 is that out of the 20 studies which found a significant, or a notable effect of gender, 10 could be interpreted to be concordant with the definition of gender, seven could not, and three could be partly interpreted in line with the definition of gender. This partly supports hypothesis three in that only 50% of the examined studies could be interpreted in line with the definition of gender. Methodology has been found to be an important factor in determining the extent to which the findings could have been interpreted as actually investigating gender, or whether they may have conflated gender with other variables, such as cognition, or sex. This conflation stems from insufficient differentiation of the concepts of *gender,* and *sex*, likely occurring because of researchers' heteronormative views of the two concepts, and thus the findings support hypothesis four.

5. Discussion

This section discusses the findings of this dissertation by considering the aspects of validity, reliability, and methodological rigour. Firstly, the importance of construct, divergent, and content validity are discussed, before focusing on the effects of questionnaires, and context on reliability. LLSs, and learning styles are discussed separately because of their inherent confounding of *gender*, and *sex*. These aspects result in calls for added methodological rigour, which includes increased awareness of gender theories, methodological consciousness, ultimately leading to calls for methodological innovation in the field.

5.1 Validity

5.1.1 Construct validity

This dissertation has found that in the systematically collected sample of current research examining gender in applied linguistics (n = 36), only one study included a definition of the concept, and only five studies stated using a valid method in which the information on the participants' gender was collected. These findings point to a potentially acute lack of construct validity in the examined body of research.

Construct validity can be defined as the presence of "evidence that corroborates the claims or arguments on the variable or concept underlying the measure(s)" (Li & Prior, 2022, p. 3). For a study to have construct validity, as Li and Prior (2022) elaborate, researchers ought to validate their measures by stating the nature of the concept, variable, or the construct under examination to ensure they are able to collect relevant data to support the examined notion. In the case of research examining gender in applied linguistics, a valid statement about the nature of gender would be its definition which should be included. By not defining gender in research examining this variable, as found to be the case in

most examined studies, the researchers risk not being able to gather relevant data that corresponds to the concept of *gender,* and potentially conflating it with a different concept, such as *sex*.

5.1.2 Divergent validity

Similarly, not including a valid method in which the information on the participants' gender was collected threatens the external validity of the studies. What constitutes a valid method of doing this could be subject to discussion in its own right. As outlined in the literature review, the conventional method of capturing information on gender on a dichotomous scale (male – female) is not necessarily valid because it threatens its divergent validity which "refers to whether the measure of a construct is different … from measures of constructs hypothesized to be different from the construct in question" (Li & Prior, 2022, p. 3). The dichotomous scale method imposes gender onto people based on their biological sex, and it, therefore, lacks in its ability to differentiate it from *sex* because of people's socio-cultural expectations about what *gender* is (i.e., that *sex* and *gender* are aligned, and that a biological man always has a male gender, and a biological woman is always female). The two-question method, or the sliding scale method, which were both discussed in the literature review, provide researchers with the opportunity to separate *gender* from *sex*, thus increasing the validity of a study.

In this dissertation, out of the five studies which were found to be stating a valid method, three required a dichotomous response, and one used a question with three options to ascertain the participants' gender. Based on the above discussion, it can be posited that the methodology used in these four studies to collect the information on the participant gender does not have sufficient levels of divergent validity. Only one study claimed to use what could be described as a valid method of capturing information on the participant's gender, and that was an open-ended question which allowed the participants to self-identify their

gender. This can be viewed as valid because this method does not categorise people on a dichotomous gender scale, and the participants are able to state their gender freely. Having said that, this needs to be critiqued because despite the authors' claims to be using said method, their appendix included a dichotomous gender option, not an open-choice question. This means that potentially none of the 36 examined studies used a valid method for collection of information on participant gender. More crucially—even if we set the concept of validity aside—these five studies are the *only* five studies out of a total of 36 which even stated the method in which they collected this information in any detail, despite four studies claiming that participants had been specifically chosen based on their gender. Such findings point to a considerable lack of divergent validity in the examined applied linguistics research.

5.1.3 Content validity

The lack of definitions, and the presence of a valid method to capture participants' gender further threatens the content validity of the research. Without due clarity on the nature of the investigated variable, the researchers cannot ascertain whether their methods are relevant, and sufficient enough to examine the concept they are claiming to be examining. This is supported by the findings of this thesis where only 50% (n = 10) of the studies on which text analysis was conducted were found to produce findings which were possible to interpret in line with the definition of gender. The second half of the examined studies could not have been interpreted in line with the definition of gender because of the methodology used – the methodology did not draw on socio-cultural expectations, interaction, nor the role of context in the participants' performance.

Such a divide between the concept, and the methods used to examine it can nullify "the alignment between the conceptualization and operationalisation of the content of the construct" (Li & Prior, 2022, p. 3), and invalidate the findings of

the studies in question. In practical terms, the studies whose findings could not have been interpreted in line with the definition of *gender* could have examined a different variable, such as *sex,* cognitive abilities, a different variable altogether, or they conflated *gender* with said other variables. This is particularly pertinent because the results of research on gender in applied linguistics are often contradictory, and inconclusive, as outlined in the literature review. The findings of this thesis indicate that it might be the lack of research validity that causes this inconclusiveness.

5.2 Reliability

5.2.1 Use of questionnaires

Related to validity, and the notion of using the right methods to investigate particular concepts, is the use of questionnaires in gender research, which ought to be discussed separately. Questionnaires are "any written instruments that present respondents with a series of questions or statements to which they are to react either by writing out their answers or selecting them among existing answers" (Brown, 2001, p. 6). They are a useful, qualitative method which can provide insight into participants' motivation, and beliefs, something which other data might not offer (Mackey & Gass, 2015). However, questionnaires are susceptible to the presence of bias (Mackey & Gass, 2015), especially in the case of sensitive, or potentially threatening questions, such as those on demographic details (Sudman & Bradburn, 1982, as cited in Cohen et al., 2018). This is pertinent to the answer to RQ2, as 13 studies were found to be using demographic questionnaires to capture the data on the participants' gender. Their use, whilst potentially valid in terms of divergence from the concept of *sex*, if phrased correctly, should be treated with caution because of the bias resulting from the sensitive information questionnaires require. This can be in the form of responding in line with the researchers' expectations, for example by stating their gender according to heteronormative expectations, rather than the gender the participant identifies as. This, henceforth, becomes a problem of reliability.

Reliability "refers to the consistency of candidates' responses to items of a given measure or instrument" (Li & Prior, 2022, p. 3). This can be negatively affected by bias which occurs either in case of acquiescence, "where respondents tend to agree with the statement being made, regardless of its content" (Krosnick & Presser, 2010, as cited in Cohen et al., 2018, p. 492), or when "respondents may also give an answer in terms of what they think is socially desirable, rather than what they really feel" (Cohen et al., 2018, p. 492). Such bias is particularly applicable to the discussion of the results of RQ3: if participants feel that they need to give certain answers which are socio-culturally expected of them, regardless of guaranteed anonymity offered by questionnaires, the instrument forthwith becomes an instrument for measuring socio-cultural expectations. The use of questionnaires, or similar, self-reported measures in the text analysis for RQ3 was noted in 12 out of 20 studies. Four of these (S16, S17, S27, S29) were interpreted not to be in line with the definition of gender because of the evident lack of socio-cultural, contextual, or interactional impact. In case of the other eight studies which used questionnaires, the impact of the three elements constituting the working definition of gender in this dissertation could have been significant enough to have potentially affected the results, and thus these studies were interpreted to be in line with the definition of gender. This impact of interactional, contextual, and socio-cultural factors can be seen as leading to increased levels of unreliability because in a different socio-cultural context, or in a different interaction, the participants could have supplied different answers, thus highlighting the presence of bias in their answers. Therefore, researchers should be cautious when using questionnaires to research the concept of *gender*: questionnaires are susceptible to an inherent bias, since gendered behaviour is entrenched in varying socio-cultural contexts.

5.2.2 Effect of context

The role of context might, in itself, have a considerable effect on the reliability of a study. To illustrate, study S17, which worked with heterogenous nationality groups, i.e., where all the participants come from different socio-cultural contexts, or backgrounds, found that males from different backgrounds reported having different learning styles. Thus, even though they reported an overall gender difference for some findings, had the study been conducted with respective homogenous nationality groups separately, their findings would have been contradictory. This affects the reliability of the study because its results would have been different depending on the context, and researchers should be aware that gender is socio-culturally, and contextually influenced to ensure their methods of researching gender are sufficiently rigorous. This might include splitting results by both gender, and socio-cultural background simultaneously, and running within-, and in-between group comparisons, and/ or analyses, as was the case in study S17.

These socio-cultural, and contextual effects can also be noticed in how gender is perceived by researchers from different contexts. The results to RQ2 further indicated that the practice of selecting participants based on their gender to ensure equal participation was noted in four studies, all of which were conducted in Iran. Whilst Iran can be viewed as progressive because of allowing sex reassignment surgery for people whose gender is not aligned with their biological sex (Pirnia & Pirnia, 2022), these beliefs stem from a dichotomous perspective of gender along the biological sex divide. As such, and based on the findings of this dissertation, it could be hypothesised that such cultural views of gender can lead to a lack of methodological rigour in researchers coming from contexts where gender might not be perceived as something to accurately describe, collect information on, and measure.

5.2.3 Language learning strategies, and learning styles

The issues of reliability, and validity also have an effect on the research of learning styles, and LLSs. LLSs encompass cognitive, and metacognitive strategies, and in the case of Oxford's (1990) SILL, which is "the most widely used instrument in language learner strategy research' (White et al. 2007, as cited in Rose, 2015, p. 425), they also include memory, compensation, affective, and social strategies. LLSs have historically been researched using "questionnaires based on inventories of reported strategy use" (Rose, 2015, p. 424) which potentially challenges both their reliability, and validity because of the self-reported nature of questionnaires (Rose, 2015) – participants need certain metacognitive strategies to be able to reflect on their use of metacognitive strategies. Rose (2015) notes that triangulating questionnaire data, and supplementing it with qualitative data could help resolve these issues. Such an approach would be useful in terms of providing the much-needed insight into the context of the studies to help ascertain the effects of socio-cultural expectations.

Relevant to this dissertation, however, is the problematic categorisation, and definition of LLSs. Rose (2015) calls LLSs "processes and actions that are consciously deployed by language learners to help them to learn or use a language more effectively" (p. 422). Cohen (2014) uses the terms "thoughts and actions" to define LLSs (p. 7), whilst Oxford (1994) states that LLSs are "behaviours" (p. 145). The term "behaviour" could be understood as something done in an interaction with other people, i.e., it could be seen to be in line with the definition of gender. However, "thoughts", and "processes" could not be seen to be performed in an interaction, and thus a question of whether LLSs are influenced by socio-cultural expectations, and whether they are performed in an interaction arises. This issue was also highlighted by Macaro (2006) who problematises "whether strategies occur inside, or outside of the brain" (p. 325), i.e., whether they are affected by socio-cultural influences, or biology. This could

be seen in the text-analysis of study S32 which utilised SILL – the breakdown of the results included strategies, such as "Managing your emotions" which, alongside the broader categories of social, and affective strategies, could be seen as being in line with the definition of gender, i.e., something that occurs in an interaction, it is context-dependent, and potentially socio-culturally influenced. What this leads this dissertation to suggest, then, is that when gender is being investigated as an independent variable alongside LLSs as the dependent variable, the results might be mixed, and inconclusive because LLSs comprise strategies which are both in line with definition of gender, as well as those which are not, thus conflating *gender* with *sex*, two concepts which are innately different.

A similar observation was made in the analysis of study S17 which investigated learning styles. Learning styles can be defined as "overall patterns that give general direction to learning behaviour" (Cornett, 1983, as cited in Oxford, 1994, p. 140), and they are seemingly related to LLSs because in a situation where "learners are not pressed to use certain strategies, they tend to use those congruent with their style" (Oxford, 1994, p. 145). The study found that the participants' background had an impact on the reported learning styles which differed amongst men from different countries. Although this, on its own, was not substantial enough to warrant interpreting the findings in line with the definition of gender because learning styles are not performed in an interaction, the presence of contextual influence does raise the question whether learning styles occur "inside or outside of the brain", to use Macaro's (2006) words. A deeper look into Oxford's (1994) overview of what comprises learning styles suggests that while some aspects are context-, and socio-culturally dependent, such as field dependence, others are innate, such as lateralisation of the brain function. Oxford (1994) herself acknowledges the gender – sex split when referring to Maccoby and Jackling (1974, as cited by Oxford, 1994): "these differences may, at least in part, be innate - and thus in fact sex differences - but

most are likely to be socio[-]culturally developed" (p. 141). The implications this has for applied linguistics research examining gender as an independent variable is that learning styles, like LLSs, confound socio-cultural, and biological factors, leading to conflicting, and inconclusive findings. The way forward is to be methodologically more rigorous in separating aspects of learning styles, and LLSs that pertain biology from those that pertain socio-cultural, interactional, and contextual factors, and researching them as influenced by two separate independent variables – *sex*, and *gender* respectively.

5.3 Methodological rigour

5.3.1 Theories on gender vs sex

Some studies—albeit a very limited number thereof—have demonstrated some awareness of gender, and an understanding of what the concept entails. In this dissertation, one study (S5) was found to conceptualise gender as a socio-cultural phenomenon. In the literature review, two studies mentioned nature-, and nurture-based theories which refer to *sex*, and *gender* respectively: Leaper and Ayres (2007), and Van Der Slik et al. (2015).

Leaper and Ayres (2007) conducted a meta-analysis of gender variations in talkativeness, affiliative, and assertive speech. In their literature review, they outline three explanations for gender differences in language use: socialisation, social constructionist/ contextualist, and biological. The socialisation explanation posits that gender-typed activities result in males and females developing different preferences, including how assertion, and affiliation are expressed through language use. The contextualist explanation is similar in that it highlights the role of interaction in context in determining how males, and females act, and respond. The biological explanation presupposes that female, and male brains are different from each other in terms of organisation, and function, thus leading to differences in language abilities, and language use (Leaper & Ayres, 2007). From these definitions it could be argued that the first

two are in line with the definition of *gender*, as they focus on interaction in socio-cultural context, whereas the biological one is related to *sex*. The very existence of such an acknowledgment in a study researching gender is noteworthy, and applied linguistics research examining gender as a variable needs this sort of awareness of what gender means.

Similar awareness of nurture-, and nature-based theories was demonstrated by Van Der Slik et al. (2015) who conducted a study in the Netherlands which compared gender differences in language acquisition amongst immigrants from 88 countries. In their discussion, the authors touch on nature-based explanations, as well as two specific nurture-based theories, namely the human capital approach, and gender-specific acculturation, both of which can be interpreted in line with the definition of gender. Although the researchers conclude that their findings "give strong circumstantial evidence of an initially nature-based gender distinction" (Van Der Slik et al., 2015, p. 18), such a distinction needs to be fundamental to the study design to ensure that the studies are clear whether they are investigating *gender*, i.e., a nurture-based factor, or *sex*, which is a nature-based concept.

5.3.2 Increasing the validity of studies/ methodological consciousness

It could be speculated that the lack of a definition of gender on its own may not necessarily mean too much, as long as the authors are aware of the differences between the socio-cultural/ contextual/ interactional factors on one hand, and the biological aspects on the other, and as long as they are referred to as such in the discussion of the findings. Having said this, however, these two concepts have labels – *gender* and *sex* respectively, and as this dissertation has found, they should not only be distinguished as such, but their disambiguation also ought to be fundamental right from the onset of research, and reflected in its methodology.

Calling for such measures might be labelled as demanding methodological rigour pertaining the examination of gender as a variable. This dissertation examined 36 studies for their definition of the term *gender*, and for the way they collected the information on their participants' gender, and analysed 20 of those studies to see if their findings could be interpreted in line with the definition of gender. Given the scope of this thesis, it is acknowledged that 36 studies are a representative sample of the existing applied linguistics research into gender, and the generalisability of the findings could be questioned; it is noted that there might have been other studies that had not been published on Scopus that do gender due methodological justice. Should the findings of this thesis be generalisable to the wider applied linguistics research, however, then the call for *methodological rigour* is not sufficient. If a research field demonstrates such potentially serious lack of methodological detail, then one should call for *methodological innovation* in the field.

Methodological innovation stems from the concept of *methodological consciousness* which "recognises an obligation to address the processes and procedures of research, not just our objects of study or our data and empirical findings" (Li & Prior, 2022, p. 2). The first step in this is having substantive domain knowledge because lack thereof "results in flawed research designs and misleading findings, which will have deleterious consequences for an applied field characterized by evidence-based practice" (Li & Prior, 2022, p. 2). This is evident from the findings pertaining to gender in this dissertation, which is why better understanding of the concept of gender, demonstrating this understanding by including its definition, and problematising it in the literature reviews of published research is an essential step that needs to be called for. This consciousness then needs to be extended by incorporating the four characteristics of methodological innovation outlined in Li et al. (2023): originality, better quality, methodological literacy, and spirt or mindset. Research into gender in applied linguistics, as demonstrated in this thesis, should be

original, and better quality, in that it should aim to "generate new evidence … by overcoming existing limitations" (Li et al., 2023, p. 551). Limitations have been shown to be present in the way applied linguistics studies capture the information on participants' gender, and in mis-conceptualising *gender* by conflating it with *sex*. Increased consciousness, and understanding of this issue will promote methodological literacy, so that this understanding is reflected in the research methodology, and that best practices are followed. These could include using the two-question method described by GenIUSS (2014), and Truman et al., (2019), or the sliding scale method used by Magliozzi et al. (2016), to name a couple of suggestions. This ought to be done to foster the right mindset amongst researchers, who should aspire to be "well-informed, open-minded, critical, progressive, adaptable, reflective, and ethical" (Li et al., 2023, p. 551).

Calls for methodological innovation in applied linguistics as a result of flawed methodology are echoed by other authors. Li (2018) synthesised methodology of feedback research, and, similarly to this thesis, found a lack of external, and internal validity in some studies. Gass and Mackey (2012), and Plonsky (2014) found study quality to be a sizable issue in the field (as cited in Li & Prior, 2022), and relevantly to gender, Sunderland (2000) highlights the two "outdated, theoretically unsophisticated concepts of gender", i.e., its heteronormative, binary categorisation, and the way it is determined for individuals, as a major problem in education research. Something ought to fundamentally change in the way the applied linguistics field researches gender as an independent variable, and it is the approach to defining, and measuring the concept in a valid, and reliable way that needs to be innovated.

6. Conclusion

6.1 Summary

This study has examined the treatment of gender in applied linguistics on a representative sample of recent research (n = 36). It has found that this treatment is largely insufficient: only one study included a definition of the concept of *gender*, and only five studies stated how they collected the information on participants' gender, whilst only one of them could potentially be considered as valid if taken at face value as stated by the researchers. These findings support hypotheses one, and two: most current research in applied linguistics investigating the concept and/ or impact of gender does not include a definition of the concept, and it does not stipulate how the information on participants' gender was collected. As some studies claimed to have chosen participants based on their gender without using a valid methodology to collect this information, this is likely to have been determined based on the perceived participants' gender identity, and heteronormative views, thus further supporting hypothesis two.

These methodological shortcomings had an impact on the findings of the studies, since half of them could not have been clearly interpreted in line with the definition of gender (n = 10), supporting hypothesis three to a certain extent. It is argued that although this dissertation worked only with a representative sample of research, its findings reflect, and explain the inconclusive nature of the research field. This provides support for hypothesis four, since the inconclusiveness stems from drawing on heteronormative views which result in aligning participants' *gender* with *sex*, two concepts which are fundamentally different, as they pertain culture/ nurture, and biology/ nature respectively, and in not approaching their investigation in a methodologically rigorous way. Although further research could investigate the sources of these heteronormative views, i.e., whether they are reflective of the researchers' personal views, or whether their expression is subconscious, their very presence

in research needs to be eliminated to ensure methodological rigour, which will subsequently create more ideal conditions to generate more valid, and reliable findings on gender as an independent variable.

6.2 Limitations

A number of factors potentially limiting the findings of this study need to be mentioned. The first is the application of the working definition of gender in this dissertation. This consisted of three elements, namely that gender is performed in an interaction, it is subject to socio-cultural expectations, and it is context specific. As this was a working definition, it had not been stipulated how many of these elements had to be met for a study to be interpreted in line with the definition. The author's best judgment—supported by systematic review methodology to reduce bias—was used to interpret the studies as objectively as possible, but on occasion, the decisions could be questioned. Examples of this would include study S17 whose results were not interpreted in line with the definition of gender, but the study did report context-specific findings, or study S15 which did not include, nor did it research interaction, but it did investigate culturally-driven behaviours. To overcome this limitation, future research of this type could prioritise the definitional elements, or it could stipulate how many need to be fulfilled to satisfy the working definition.

Secondly, it is acknowledged that the author is not a specialist in any particular niche of the applied linguistics research field. As the study covers a wide range of research areas, and dependent variables, it was beyond the scope of this study to acquire sufficiently comprehensive understanding of all variables to be able to make truly informed decisions. This broad range of studies covered is, however, a representative sample only. While there are other studies known to the author which did fulfil the inclusion criteria, a systematic approach to study collection had been chosen to avoid researcher bias. It is believed that the representative sample was sufficient to draw conclusions about the state of

applied linguistics research pertaining to gender, but it needs to be acknowledged that there potentially—and hopefully—are other studies in the field which treat gender more rigorously.

It also needs to be noted that AntFile Converter (Anthony, 2022) used to prepare studies for corpus analysis occasionally added spaces between letters. The impact of this was mitigated by manually scanning the studies, but omissions could have occurred in the process. Despite focusing on adults, some studies did include a proportion of underage participants, and the chosen cut-off point of 2004 used for the inclusion criteria could be considered arbitrary, and not representative of the developments in the treatment of gender in the wider, international context.

6.3 Recommendations for future research

The difference between *gender*, and *sex* is important to transgender people. Despite the findings of this study potentially having some implications for this demographic, their main implication is on methodological rigour in applied linguistics, and social science research. As the two concepts pertain to culture, and biology respectively, methodological approaches researching them also ought to be examining either cultural, or biological aspects of human beings to ensure that the studies are valid. As discussed in this dissertation, based on the current state of research, methodological *innovation*, not just rigour, needs to be called for. This can be achieved via a number of recommendations.

The first recommendation is to ensure that researchers have demonstratable knowledge of the concept of gender. This ought to be reflected in the literature reviews, and methodologies of studies examining gender as a variable by including a problematisation of the concept, or at least its definition. This forms part of methodological consciousness without which the relevance of the

methods of a study will be limited (Li & Prior, 2022) because lack thereof would not demonstrate a sufficient understanding of the concept under investigation.

Subsequently, this consciousness ought to be reflected in the methodology of a study in two different ways. Firstly, researchers need to ensure they collect the information on participants' gender in a valid, and reliable way, i.e., to ensure divergent validity, so that *gender* can be distinguished from *sex*. Suggestions mentioned in this study include the two-question method described by GenIUSS (2014), and Truman et al., (2019), or the sliding scale method used by Magliozzi et al. (2016). Secondly, the method of investigating the dependent variable needs to ensure content validity. This can be achieved by using mixed-methods to supplement quantitative data with qualitative insight (Rose, 2015; Sunderland, 2000), especially in case of longitudinal studies, or studies which use delayed post-tests. The qualitative insight would take the form of capturing the presence of socio-cultural expectations, which would also be useful when researching LLSs which ought to be separated into those pertaining to behaviours, which can be socio-culturally influenced and/ or performed in interactions, and biology/ cognition. This would, as a result, help "untie the Gordian knot of style and strategy differences" (Oxford, 1994, p. 147) stemming from the lack of disambiguation between learner's *sex*, and *gender*, as Oxford (1994) calls for. Researchers also ought to be conscious of the socio-cultural bias questionnaires can lead to, and use them knowing that their use will inevitably lead to researching an element of culture/ gender. Lastly, attention should be paid to context, particularly cultural homogeneity of research groups: since gender is a part of culture, culturally heterogenous groups are unsuitable for researching the impact of gender on the dependent variable in question because participants in such groups could be seen to have multiple different genders intertwined with their culture.

As Cohen et al. (2018) assert, “methodological rigour is an ethical, not simply a technical matter, and respondents have a right to expect reliability and validity” (p. 472). This study has identified such an acute lack of methodological rigour in applied linguistics research on gender that it calls for methodological innovation to ensure reliability, and validity. It is hoped that this dissertation has not only highlighted an issue in the research field, but also offered a number of useful solutions to overcome the situation. It is, after all, the purpose of research to investigate, inform, and provide a way forward in its field.

References

Adel, S., & Enayat, M. (2016). Gender representation and stereotyping in ESP textbooks. *Asian ESP Journal*, *12*(3), 94–119.

Ansary, H., & Babaii, E. (2003). On the manifestation of subliminal sexism in current Iranian secondary school ELT textbooks. *Iranian Journal of Applied Linguistics*, *6*(1), 40–56.

Anthony, L. (2022). *AntFileConverter* (Version 2.0.2). [Computer Software]. Waseda University. Available from https://www.laurenceanthony.net/software

Anthony, L. (2023). *AntConc* (Version 4.2.4). [Computer Software]. Waseda University. Available from https://www.laurenceanthony.net/software

Aydınoğlu, N. (2014). Gender in English language teaching coursebooks. *Procedia - Social and Behavioral Sciences*, *158*, 233–239. https://doi.org/10.1016/j.sbspro.2014.12.081

Bacon, S. M. (1992). The relationship between gender, comprehension, processing strategies, and cognitive and affective response in foreign-language listening. *The Modern Language Journal*, *76*(2), 160–178. https://doi.org/10.2307/329769

Bacon, S. M. C., & Finnemann, M. D. (1992). Sex differences in self-reported beliefs about foreign-language learning and authentic oral and written input [Article]. *Language Learning*, *42*(4), 471–495. https://doi.org/10.1111/j.1467-1770.1992.tb01041.x

Badger, D., Nursten, J., Williams, P., & Woodward, M. (2000). Should all literature reviews be systematic? *Evaluation & Research in Education*, *14*(3–4), 220–230. https://doi.org/10.1080/09500790008666974

Barton, A., & Sakwa, L. N. (2012). The representation of gender in English textbooks in Uganda. *Pedagogy, Culture & Society*, *20*(2), 173–190. https://doi.org/10.1080/14681366.2012.669394

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, *8*(4), 243–257. https://doi.org/10.1093/llc/8.4.243

Blumberg, R. L. (2008). *Gender bias in textbooks: A hidden obstacle on the road to gender equality in education*.

Boyle, J. P. (1987). Sex differences in listening vocabulary. *Language Learning*, *37*(2), 273–284. https://doi.org/10.1111/j.1467-1770.1987.tb00568.x

Brown, J. D. (2001). *Using surveys in language programs*. Cambridge University Press.

Butler, J. (2004). *Undoing gender*. Routledge. https://doi.org/10.4324/9780203499627

Butler, J. (2006). *Gender trouble*. Routledge. https://doi.org/10.4324/9780203824979

Cambridge University Press and Assessment. (n.d.). Gender. In *Cambridge Dictionary*. Retrieved August 27, 2023, from https://dictionary.cambridge.org/dictionary/english/gender

Chalmers, H. (2023). *International Database of Education Systematic Reviews*. Https://Idesr.Org.
Chavez, M. (2000). Teacher and student gender and peer group gender composition in German foreign language classroom discourse: An exploratory study. *Journal of Pragmatics*, *32*(7), 1019–1058. https://doi.org/10.1016/S0378-2166(99)00065-X
Cohen, A. D. (2014). *Strategies in learning and using a second language* (2nd ed.). Routledge. https://doi.org/10.4324/9781315833200
Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.
Colllins. (n.d.). Sex. In *Collins COBUILD Advanced Learner's Dictionary*. Retrieved June 13, 2024, from https://www.collinsdictionary.com/dictionary/english/sex
Coxhead, A. (2020). Analysis of corpora. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 464–473). Routledge.
Curthoys, A. (2014). Gender in the social sciences. *Australian Feminist Studies*, *29*(80), 115–120. https://doi.org/10.1080/08164649.2014.930553
Derakhshan, A., Malmir, A., Pawlak, M., & Wang, Y. (2023). The use of interlanguage pragmatic learning strategies (IPLS) by L2 learners: The impact of age, gender, language learning experience, and L2 proficiency levels. *IRAL - International Review of Applied Linguistics in Language Teaching*. https://doi.org/10.1515/iral-2022-0132
Ehrman, M., & Oxford, R. L. (1989). Effects of sex differences, career choice, and psychological type on adult language learning strategies. *The Modern Language Journal*, *73*(1), 1–13. https://doi.org/10.1111/j.1540-4781.1989.tb05302.x
Ekstrand, L. H. (1980). Sex differences in second language learning? Empirical studies and a discussion of related findings. *International Review of Applied Psychology*, *29*(1), 205–259.
Fausto-Sterling, A. (1992). *Myths of gender: biological theories about women and men* (2nd ed.). Basic Books.
Fausto-Sterling, A. (2000). *Sexing the body: Gender politics and the construction of sexuality*. Basic Books.
Geddes, P. (1895). *The evolution of sex. By Professor Patrick Geddes and J. Arthur Thompson. With 104 illustrations*. W. Scott, C. Scribner, 1895.
GenIUSS. (2014). *Best practices for asking questions to identify transgender and other gender minority respondents on population-based surveys*. eScholarship, University of California.
Gordon, D. (2004). "I'm tired. You clean and cook." Shifting gender identities and second language socialization. *TESOL Quarterly*, *38*(3), 437–457. https://doi.org/10.2307/3588348
Gorman, C., & Nash, J. M. (1992). Sizing up the sexes. *Time (Chicago, Ill.), 139*(3), 42.

Goyal, R., & Rose, H. (2020). Stilettoed Damsels in Distress: The (un)changing depictions of gender in a business English textbook. *Linguistics and Education*, *58*, 100820–100829. https://doi.org/10.1016/j.linged.2020.100820

Gray, J. (2000). The ELT coursebook as cultural artefact: How teachers censor and adapt. *ELT Journal*, *54*(3), 274–283. https://doi.org/10.1093/elt/54.3.274

Hall, J. A. (1984). *Nonverbal sex differences: Communication accuracy and expressive style*. The Johns Hopkins University Press.

Haslanger, S. (2000). Gender and race: (What) are they? (What) do we want them to be? *Noûs (Bloomington, Indiana)*, *34*(1), 31–55. https://doi.org/10.1111/0029-4624.00201

Hilliard, A. D. (2014). A critical examination of representation and culture in four English language textbooks. *Language Education in Asia*, *5*(2), 238–252. https://doi.org/10.5746/LEiA/14/V5/I2/A06/Hilliard

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, *104*(1), 53–69. https://doi.org/10.1037/0033-2909.104.1.53

International Olympic Committee. (2004). *IOC approves consensus with regard to athletes who have changed sex*. Consensus Statements. https://olympics.com/ioc/documents/athletes/medical-and-scientific-consensus-statements

Jaggar, A. (1983). Human biology in feminist theory: Sexual equality reconsidered. In C. Gould (Ed.), *Beyond domination: New perspectives on women and philosophy* (pp. 21–42). Rowman & Littlefield Publishers, Inc.

James, D., & Drakich, J. (1993). Understanding gender differences in amount of talk: A critical review of the research. In D. Tannen (Ed.), *Gender and conversational interaction* (pp. 281–312). Oxford University Press.

Jasmani, M. F. I. M., Yasin, M. S. M., Hamid, B. A., Keong, Y. C., Othman, Z., & Jaludin, A. (2011). Verbs and gender: The hidden agenda of a multicultural society. *Journal of Language Teaching, Linguistics, and Literature*, *17*(special issue), 61–73.

Khaneshan, P. Y., & Bonyadi, A. (2016). The investigation of compliment response patterns across gender and age among advanced EFL learners. *Journal of Language Teaching and Research*, *7*(4), 760–767. https://doi.org/10.17507/jltr.0704.17

Koul, R., Roy, L., Kaewkuekool, S., & Ploisawaschai, S. (2009). Multiple goal orientations and foreign language anxiety. *System (Linköping)*, *37*(4), 676–688. https://doi.org/10.1016/j.system.2009.09.011

Leaper, C., Anderson, K. J., & Sanders, P. (1998). Moderators of gender effects on parents' talk to their children: A meta-analysis. *Developmental Psychology*, *34*(1), 3–27. https://doi.org/10.1037/0012-1649.34.1.3

Leaper, C., & Ayres, M. M. (2007). A meta-analytic review of gender variations in adults' language use: talkativeness, affiliative speech, and assertive speech. *Personality and Social Psychology Review*, *11*(4), 328–363. https://doi.org/10.1177/1088868307302221

Leaper, C., & Smith, T. E. (2004). A meta-analytic review of gender variations in children's language use: Talkativeness, affiliative speech, and assertive speech. *Developmental Psychology*, *40*(6), 993–1027. https://doi.org/10.1037/0012-1649.40.6.993

Lee, J. F. K. (2014). A hidden curriculum in Japanese EFL textbooks: Gender representation. *Linguistics and Education*, *27*, 39–53. https://doi.org/https://doi.org/10.1016/j.linged.2014.07.002

Lewandowski, M. (2014). Gender stereotyping in EFL grammar textbooks: A diachronic approach. *Linguistik Online*, *68*(6), 83–99. https://doi.org/10.13092/lo.68.1635

Li, S. (2018). Data collection in the research on the effectiveness of corrective feedback: A synthetic and critical review. In A. Gudmestad, & A. Edmonds (Eds.), *Critical reflections on data in second language acquisition* (pp. 33-61). John Benjamins.

Li, S., Prior, M., Nero, S., Hiver, P., Al-Hoorie, A. H., Murakami, A., Wei, L., & Ortega, L. (2023). Methodological innovation in applied linguistics research: Perspectives, strategies, and trends. *Language Teaching*, *56*(4), 551–556. https://doi.org/10.1017/S026144482300023X

Li, S., & Prior, M. T. (2022). Research methods in applied linguistics: A methodological imperative. *Research Methods in Applied Linguistics*, *1*(1), 100008. https://doi.org/10.1016/j.rmal.2022.100008

Liang, H.-Y., & Kelsen, B. A. (2017). Gender differences in university EFL students' language proficiency corresponding to self-rated attention, hyperactivity and impulsivity. *Electronic Journal of Research in Educational Psychology*, *15*(1), 48–74. https://doi.org/10.14204/ejrep.41.16017

Lin, L.-F. (2011). Gender differences in L2 comprehension and vocabulary learning in the video-based CALL program. *Journal of Language Teaching and Research*, *2*(2), 295–301. https://doi.org/10.4304/jltr.2.2.295-301

Liyanage, I., & Bartlett, B. J. (2012). Gender and language learning strategies: Looking beyond the categories. *Language Learning Journal*, *40*(2), 237–253. https://doi.org/10.1080/09571736.2011.574818

Ludwig, J. (1983). Attitudes and expectations: A profile of female and male students of college French, German, and Spanish. *The Modern Language Journal*, *67*(3), 216–227. https://doi.org/10.1111/j.1540-4781.1983.tb01499.x

Macaro, E. (2006). Strategies for language learning and for language use: Revising the theoretical framework. *The Modern Language Journal*, *90*(3), 320–337.

Macaro, E. (2020). Systematic reviews in applied linguistics. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 230–239). Routledge.

Mackey, A., & Gass, S. M. (2015). *Second language research: Methodology and design* (S. M. Gass, Ed.; 2nd ed.). Routledge, an imprint of Taylor and Francis. https://doi.org/10.4324/9781315750606

Magliozzi, D., Saperstein, A., & Westbrook, L. (2016). Scaling up: Representing gender diversity in survey research. *SAGE*, *2*, 1–11. https://doi.org/10.1177/2378023116664352

Mineshima, M. (2008). Gender representation in an EFL textbook. *Bulletin of Nigata Institute of Technology*, 121–140.

Morris, L. A. (1998). Differences in men's and women's ESL writing at the junior college level: Consequences for research on feedback. *Canadian Modern Language Review*, *55*(2), 216–238. https://doi.org/10.3138/cmlr.55.2.219

Nation, I. S. P. (2016). *Making and using word lists for language learning and testing* [Book]. John Benjamins Publishing Company.

Newman, M., & Gough, D. (2020). Systematic reviews in educational research: Methodology, perspectives and application. In O. Zawacki-Richter, M. Kerres, S. Bedenlier, M. Bond, & K. Buntins (Eds.), *Systematic reviews in educational research: Methodology, perspectives and application* (pp. 3–22). Springer VS.

Nicholson, L. (1994). Interpreting gender. *Signs: Journal of Women in Culture and Society*, *20*(1), 79–105. https://doi.org/10.1086/494955

Ogbay, S. (1999). *The social and linguistic construction and maintenance of girls' and boys' gender identity in two secondary schools in Eritrea*. Lancaster University.

Otlowski, M. (2003). Ethnic diversity and gender bias in EFL textbooks. *Asian EFL Journal*, *5*(2), 1–15.

Oxford, R. (1994). La différence continue…: Gender differences in second/foreign language learning styles and strategies. In J. Sunderland (Ed.), *Exploring gender: Questions and implications for English language education* (pp. 140–147). Prentice Hall.

Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Heinle & Heinle.

Oxford, R., & Nyikos, M. (1989). Variables affecting choice of language learning strategies by university students. *The Modern Language Journal*, *73*(3), 291–300. https://doi.org/10.2307/327003

Oxford University Press. (n.d.). Gender. In *Oxford Advanced Learner's Dictionary*. Retrieved August 27, 2023, from https://www.oxfordlearnersdictionaries.com/definition/english/gender?q=gender

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ (Online)*, *372*, n71. https://doi.org/10.1136/bmj.n71

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Blackwell Publishing. https://doi.org/10.1002/9780470754887

Pichette, F., Béland, S., & Leśniewska, J. (2019). Detection of gender-biased items in the peabody picture vocabulary test. *Languages*, *4*(2). https://doi.org/10.3390/languages4020027

Pirnia, B., & Pirnia, K. (2022). Sex reassignment surgery in Iran, re-birth or human rights violations against transgender people? *Iran J Public Health*, *51*(11), 2632–2633.

Richards, J. C. (2013). Gender. In *Longman dictionary of language teaching and applied linguistics* (Richard. Schmidt, Ed.; 4th ed., p. 240). Routledge. https://doi.org/10.4324/9781315833835

Robinson, J. P., & Lubienski, S. T. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings. *American Educational Research Journal*, *48*(2), 268–302. https://doi.org/10.3102/0002831210372249

Rogers, L. J. (2000). *Sexing the brain*. Phoenix.

Rose, H. (2015). Researching language learner strategies. In B. Paltridge & A. Phakiti (Eds.), *Research methods in applied linguistics: A practical resource* (2nd ed., pp. 421–438). Bloomsbury.

Rose, H., McKinley, J., & Briggs Baffoe-Djan, J. (2019). *Data collection research methods in applied linguistics*. Bloomsbury Publishing Plc. https://doi.org/10.5040/9781350025875

Rubin, G. (1975). The traffic in women: Notes on the 'political economy' of sex. In R. Reiter (Ed.), *Toward an anthropology of women* (pp. 157–210). Monthly Review Press.

Sandland, R. (2005). Feminism and the gender recognition act 2004. *Feminist Legal Studies*, *13*(1), 43–66. https://doi.org/10.1007/s10691-005-1456-3

Schilt, K., & Westbrook, L. (2009). Doing gender, doing heteronormativity: "Gender normals," transgender people, and the social maintenance of heterosexuality. *Gender & Society*, *23*(4), 440–464. https://doi.org/10.1177/0891243209340034

Şeker, M., & Dinçer, A. (2014). An analysis of gender stereotyping in English teaching course books. *Çukurova University Faculty of Education Journal*, *43*(1), 90–98. https://doi.org/10.14812/cufej.2014.007

Sinclair, J. (2004a). Corpus and text: Basic principles. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (n.p.). Oxbow Books.

Sinclair, J. (2004b). Appendix: How to build a corpus. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (n.p.). Oxbow Books.

Söylemez, A. S. (2010). A study on how social gender identity is constructed in EFL coursebooks. *Procedia - Social and Behavioral Sciences*, *9*, 747–752. https://doi.org/10.1016/j.sbspro.2010.12.228

Spelman, E. V. (1988). *Inessential woman: Problems of exclusion in feminist thought*. Beacon Press.

Stacey, J., & Thorne, B. (1985). The Missing Feminist Revolution in Sociology. *Social Problems (Berkeley, Calif.)*, *32*(4), 301–316. https://doi.org/10.1525/sp.1985.32.4.03a00010

Stoljar, N. (1995). Essence, identity, and the concept of woman. *Philosophical Topics*, *23*(2), 261–293. https://doi.org/10.5840/philtopics19952328

Stoller, R. J. (1968). *Sex and gender: On the development of masculinity and femininity* [Book]. Science House.

Sun, C.-F., Xie, H., Metsutnan, V., Draeger, J. H., Lin, Y., & Kablinger, A. S. (2022). 3.114 The mean age of gender dysphoria diagnosis is decreasing. *Journal of the American Academy of Child and Adolescent Psychiatry*, *61*(10), S265–S265. https://doi.org/10.1016/j.jaac.2022.09.392

Sunderland, J. (2000). Research into gender and language education: Lingering problems and new directions. *The Language Teacher*, *24*(7), 8–10.

Tabler, J., Snyder, J. A., Schmitz, R. M., Geist, C., & Gonzales, C. M. (2023). Embracing complexity: Variation in faculty's attitudes toward inclusive measures of gender and sexuality in social and health sciences research. *Journal of Homosexuality*, *70*(10), 2253–2275. https://doi.org/10.1080/00918369.2022.2059967

Taylor, F., & Marsden, E. J. (2014). Perceptions, attitudes, and choosing to study foreign languages in England: An experimental intervention. *The Modern Language Journal (Boulder, Colo.)*, *98*(4), 902–920. https://doi.org/10.1111/modl.12146

Tenney, J. W., Paiva, M., & Wang, Q. (2020). Assessment of English language performance scores and academic performance in an English-based curriculum for pharmacy students with English as a second language. *Currents in Pharmacy Teaching and Learning*, *12*(4), 423–428. https://doi.org/10.1016/j.cptl.2019.12.029

Tercanlioglu, L. (2004). Exploring gender effect on adult foreign language learning strategies. *Issues in Educational Research*, *14*(2), 181–193. https://www.scopus.com/inward/record.uri?eid=2-s2.0-8744311710&partnerID=40&md5=5edb99393832bb56a13f3d719ca73fc3

Thomas, N., Bowen, N. E. J. A., & Rose, H. (2021). A diachronic analysis of explicit definitions and implicit conceptualizations of language learning strategies. *System*, *103*, 102619.

Truman, J. L., Morgan, R. E., Gilbert, T., & Vaghela, P. (2019). Measuring sexual orientation and gender identity in the National Crime Victimization Survey. *Journal of Official Statistics*, *35*(4), 835–858. https://doi.org/10.2478/jos-2019-0035

Van Der Slik, F. W. P., Van Hout, R. W. N. M., & Schepens, J. J. (2015). The gender gap in second language acquisition: Gender differences in the acquisition of Dutch among immigrants from 88 countries with 49 mother tongues. *PloS One*, *10*(11), e0142056–e0142056. https://doi.org/10.1371/journal.pone.0142056

Victor, L. (2008). Systematic reviewing. *Social Research Update*, *54*, 1–4.

Wang, W. (2020). Text analysis. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 453–463). Routledge.

West, C., & Zimmerman, D. H. (1987). Doing gender. *Gender & Society*, *1*(2), 125–151. https://doi.org/10.1177/0891243287001002002

Yang, C. C. R. (2012). Is gender stereotyping still an issue? An analysis of a Hong Kong primary English textbook series. *Hong Kong Journal of Applied Linguistics*, *13*(2), 32–48.

Yilmaz, E. (2012). *Gender representations in ELT coursebooks: A comparative study*. Middle East Technical University.

Young, D. J., & Oxford, R. (1997). A gender-related analysis of strategies used to process written input in the native language and a foreign language. *Applied Language Learning*, *8*(1), 26–43.

Zhang, X., & Ardasheva, Y. (2019). Sources of college EFL learners' self-efficacy in the English public speaking domain. *English for Specific Purposes*, *53*, 47–59. https://doi.org/10.1016/j.esp.2018.09.004

Zittleman, K., & Sadker, D. (2002). Gender bias in teacher education texts. *Journal of Teacher Education*, *53*(2), 168–180. https://doi.org/10.1177/0022487102053002008

Appendix 1: References of included studies

Barrios, E., & Acosta-Manzano, I. (2021). Factors predicting classroom WTC in English and French as foreign languages among adult learners in Spain. *Language Teaching Research*, 136216882110540. https://doi.org/10.1177/13621688211054046

Chacón-Beltrán, R., & Echitchi, R. (2021). Who wants to learn English online for free? *Journal of Language and Education*, *7*(4), 53–65. https://doi.org/10.17323/JLE.2021.11906

Chiang, H.-H. (2020). Kahoot! in an EFL reading class. *Journal of Language Teaching and Research*, *11*(1), 33–44. https://doi.org/10.17507/jltr.1101.05

Davydova, J. (2022). The role of social factors in the acquisition of vernacular English: A variationist study with pedagogical implications. *International Journal of Applied Linguistics*, *32*(3), 425–441. https://doi.org/10.1111/ijal.12438

Derakhshan, A., Malmir, A., Pawlak, M., & Wang, Y. (2023). The use of interlanguage pragmatic learning strategies (IPLS) by L2 learners: The impact of age, gender, language learning experience, and L2 proficiency levels. *IRAL - International Review of Applied Linguistics in Language Teaching*. https://doi.org/10.1515/iral-2022-0132

Es-Hagi Sardroud, S. J. (2013). Impact of training deep vocabulary learning strategies on vocabulary retention of Iranian EFL learners. *International Journal of Applied Linguistics and English Literature*, *2*(3), 75–82. https://doi.org/10.7575/aiac.ijalel.v.2n.3p.75

Garcia de Blakeley, M., Ford, R., & Casey, L. (2017). Second language anxiety among Latino American immigrants in Australia. *International Journal of Bilingual Education and Bilingualism*, *20*(7), 759–772. https://doi.org/10.1080/13670050.2015.1083533

Ghasemi, F., Mohammadnia, Z., & Gholami, Z. (2023). Individual differences in teacher hopelessness: Examining the significance of personal and professional factors. *Psychology Hub*, *40*(2), 49–58. https://doi.org/10.13133/2724-2943/17816

Gordon, D. (2004). "I'm tired. You clean and cook." Shifting gender identities and second language socialization. *TESOL Quarterly*, *38*(3), 437–457. https://doi.org/10.2307/3588348

İlya, A. (2023). Factors affecting Turkish EFL teachers' level of burnout: A quantitative study. *Journal of Pedagogical Research*, *7*(1), 142–153. https://doi.org/10.33902/JPR.202317925

Inada, T. (2021). Are there gender differences in anxiety and any other factors among English as a foreign language (EFL) college students in Japan? *International Medical Journal*, *28*(1), 90–93. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85118103318&partnerID=40&md5=67ff0c177736f679f24ea9b9414bf48c

Ju, C. (2009). In-service adult learners' English learning strategies in Taiwan. *International Journal of Learning*, *16*(10), 119–132. https://doi.org/10.18848/1447-9494/cgp/v16i10/46653
Kentmen, H., Debreli, E., & Yavuz, M. A. (2023). Assessing tertiary Turkish EFL learners' pragmatic competence regarding speech acts and conversational implicatures. *Sustainability (Switzerland)*, *15*(4). https://doi.org/10.3390/su15043800
Khaneshan, P. Y., & Bonyadi, A. (2016). The investigation of compliment response patterns across gender and age among advanced EFL learners. *Journal of Language Teaching and Research*, *7*(4), 760–767. https://doi.org/10.17507/jltr.0704.17
Kutuk, G., Putwain, D. W., Kaye, L. K., & Garrett, B. (2022). Relations between gender stereotyping and foreign language attainment: The mediating role of language learners' anxiety and self-efficacy. *British Journal of Educational Psychology*, *92*(1), 212–235. https://doi.org/10.1111/bjep.12446
Liang, H.-Y., & Kelsen, B. A. (2017). Gender differences in university EFL students' language proficiency corresponding to self-rated attention, hyperactivity and impulsivity. *Electronic Journal of Research in Educational Psychology*, *15*(41), 48–74. https://doi.org/10.25115/EJREP.41.16017
Lincoln, F., & Rademacher, B. (2006). Learning styles of ESL students in community colleges. *Community College Journal of Research and Practice*, *30*(5–6), 485–500. https://doi.org/10.1080/10668920500207965
Minavandchal, A., & Salimi, M. (2021). Predicting EFL learners' susceptibility to various disfluency types based on gender and age. *Psiholìngvìstika*, *30*(2), 174–198. https://doi.org/10.31470/2309-1797-2021-30-2-174-198
Muñoz, C. (2020). Boys like games and girls like movies: Age and gender differences in out-of-school contact with English. *Revista Espanola de Linguistica Aplicada*, *33*(1), 172–202. https://doi.org/10.1075/resla.18042.mun
Natividad, M. R. A., & Batang, B. L. (2018). Students' perceptual learning styles and attitudes toward communicative language teaching. *TESOL International Journal*, *13*(4), 1–17. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85084244168&partnerID=40&md5=74f72087a30eca57120607bbaf376718
Nikoopour, J., Moakhar, R. K., & Esfandiari, N. (2017). The impact of explicit integrated strategies instruction on IELTS applicants' listening comprehension. *Journal of Language Teaching and Research*, *8*(4), 774–781. https://doi.org/10.17507/jltr.0804.18
Pichette, F., Béland, S., & Leśniewska, J. (2019). Detection of gender-biased items in the peabody picture vocabulary test. *Languages*, *4*(2). https://doi.org/10.3390/languages4020027
Reigel, D. (2008). Positive feedback in pairwork and its association with ESL course level promotion. *TESOL Quarterly*, *42*(1), 79–98. https://doi.org/10.1002/j.1545-7249.2008.tb00208.x

Renani, G. A., Afghari, A., & Hadian, B. (2019). Effect of awareness of teacher education philosophy on EFL teachers' professional knowledge: A postmethod perspectivisation. *International Journal of Instruction*, *12*(2), 435–454. https://doi.org/10.29333/iji.2019.12228a

Rezvani, E., & Tavakoli, M. (2013). Investigating Iranian test-takers' cognitive and metacognitive strategy use: IELTS reading section in focus. *Middle East Journal of Scientific Research*, *13*(7), 956–962. https://doi.org/10.5829/idosi.mejsr.2013.13.7.3174

Sadeghi, K., & Sharifi, F. (2013). The effect of post-teaching activity type on vocabulary learning of elementary EFL learners. *English Language Teaching*, *6*(11), 65–76. https://doi.org/10.5539/elt.v6n11p65

Shang, H.-F. (2016). Exploring demographic and motivational factors associated with hypertext reading by English as a foreign language (EFL) students. *Behaviour and Information Technology*, *35*(7), 559–571. https://doi.org/10.1080/0144929X.2015.1094827

Soltanian, N., & Sadeghi, A. (2021). Thin-slice judgments of English language teacher success in instruction: The effects of learners' gender, age, and language proficiency. *MEXTESOL Journal*, *45*(1). https://www.scopus.com/inward/record.uri?eid=2-s2.0-85103600129&partnerID=40&md5=23ca366036f9f35bc72252a986612b99

Tabatabaei, S. M., Homayoun, M. R., & Mansoorian, S. M. A. (2020). A psychoanalytic health survey on the Iranian EFL learners' emotional intelligence and their language learning approach. *International Journal of Pharmaceutical Research*, *12*(1), 733–744. https://doi.org/10.31838/ijpr/2020.12.01.144

Tavakoli, E., & Davoudi, M. (2017). Willingness to communicate orally: The case of Iranian EFL learners. *Journal of Psycholinguistic Research*, *46*(6), 1509–1527. https://doi.org/10.1007/s10936-017-9504-0

Tenney, J. W., Paiva, M., & Wang, Q. (2020). Assessment of English language performance scores and academic performance in an English-based curriculum for pharmacy students with English as a second language. *Currents in Pharmacy Teaching and Learning*, *12*(4), 423–428. https://doi.org/10.1016/j.cptl.2019.12.029

Tercanlioglu, L. (2004). Exploring gender effect on adult foreign language learning strategies. *Issues in Educational Research*, *14*(2), 181–193. https://www.scopus.com/inward/record.uri?eid=2-s2.0-8744311710&partnerID=40&md5=5edb99393832bb56a13f3d719ca73fc3

Tian, X., Samat, N. A., & Zainal, Z. (2022). Chinese EFL learners' attitudes towards smartphone-based reading. *Theory and Practice in Language Studies*, *12*(9), 1838–1847. https://doi.org/10.17507/tpls.1209.17

Weger, H. (2013). International students' attitudes toward L2-English classroom activities and language skills in the USA. *Innovation in Language Learning and Teaching*, *7*(2), 139–157. https://doi.org/10.1080/17501229.2012.733007

Xue, J. (2021). The developmental trajectory of biliteracy for Chinese–English adult EFL learners: A longitudinal study. *Reading and Writing*, *34*(4), 1089–1114. https://doi.org/10.1007/s11145-020-10105-6

Zhang, X., & Ardasheva, Y. (2019). Sources of college EFL learners' self-efficacy in the English public speaking domain. *English for Specific Purposes*, *53*, 47–59. https://doi.org/10.1016/j.esp.2018.09.004

Appendix 2: Data extraction table

Study number	S1
Authors and year of publication	Barrios & Acosta-Manzano (2021)
Title	Factors predicting classroom WTC in English and French as foreign languages among adult learners in Spain
Research question(s)	RQ1: What learner variables are associated to WTC [willingness to communicate] in adult FL learners of English and of French in a Spanish context? RQ2: How differently do learner variables predict the WTC of adult learners in English and in French as foreign languages?
Participants	420 learners of English and French FL (Spanish L1): 298 EFL learners (30.54% males, 69.49% females) 122 FFL learners (27.87% males, 72.12% females)
Setting	Language schools in Spain
Methodology	A questionnaire
Gender as a variable	Association and prediction of WTC
Method of collecting information on gender	A questionnaire Detail not provided
Results (gender)	Gender is not associated with WTC (p = 0.107)

Study number	S2
Authors and year of publication	Chacón-Beltrán & Echitchi (2021)
Title	Who wants to learn English online for free?
Research question(s)	RQ1: Are learner demographics in our courses different from other MOOCs'? RQ2: Do learner demographics vary depending on course level? RQ3: What was the impact of Covid-19 on learner demographics?
Participants	29,883 EFL learners (mostly Spanish L1 from Spanish and South American countries) (gender mix not stated)
Setting	An online course delivered by a university in Spain
Methodology	A questionnaire
Gender as a variable	Which gender is more likely to enrol in Language Massive Open Online Courses

Method of collecting information on gender	A questionnaire Three choices: “female”, “male”, “other”
Results (gender)	Female learners are more likely to take the courses than males

Study number	S3
Authors and year of publication	Chiang (2020)
Title	Kahoot! in an EFL reading class
Research question(s)	RQ1: What are EFL learners’ general perceptions of Kahoot!? RQ2: Are there any gender differences in the EFL learners’ perceptions of Kahoot!? RQ3: What advantages and disadvantages do EFL students perceive regarding the use of Kahoot! as a testing tool in the classroom?
Participants	65 EFL college students (14 males, 46 females) *Numbers do not add up
Setting	A university in Taiwan
Methodology	A questionnaire (Likert scale + 2 open questions)
Gender as a variable	Effect of gender on the perceptions of Kahoot
Method of collecting information on gender	Not stated
Results (gender)	No differences between genders were found

Study number	S4
Authors and year of publication	Davydova (2022)
Title	The role of social factors in the acquisition of vernacular English: A variationist study with pedagogical implications
Research question(s)	Not stated “The study reported here explores the role of various social factors in the L2 acquisition of vernacular variation.” p. 426
Participants	37 participants who speak English as an L2 (18 males, 19 females) General population – high school students, university students, professionals
Setting	Mannheim, Germany
Methodology	Semi-structured interviews + a background questionnaire

Gender as a variable	An independent, language-internal variable Effects of gender on the acquisition of L2 vernacular variation
Method of collecting information on gender	A questionnaire 2 options: male and female
Results (gender)	Females use innovative variants more readily (factor weight FW = 0.64; $p < 0.0349$)

Study number	S5
Authors and year of publication	Derakhshan et al. (2023)
Title	The use of interlanguage pragmatic learning strategies (IPLS) by L2 learners: the impact of age, gender, language learning experience, and L2 proficiency levels
Research question(s)	RQ1: Do L2 learners' individual differences (including age and gender) have any significant effect on the reported use of the IPLS? RQ2: Does the L2 learners' language learning experience have any significant impact on the reported use of the IPLS? RQ3: Does the L2 learners' language proficiency have any significant impact on the reported use of the IPLS?
Participants	160 EFL learners (49 males, 111 females) High school and university students
Setting	Branches of a language institute in Iran
Methodology	The Michigan Test of English Language Proficiency + an interlanguage pragmatic learning strategies (IPLS) inventory
Gender as a variable	Effects of gender on the use of interlanguage pragmatic learning strategies
Method of collecting information on gender	A questionnaire (incorporated in the IPLS inventory) Detail not provided
Results (gender)	No significant differences between male and female learners

Study number	S6
Authors and year of publication	Es-hagi Sardroud (2013)

Title	Impact of training deep vocabulary learning strategies on vocabulary retention of Iranian EFL learners
Research question(s)	RQ1: Is there any significant relationship between the explicit instruction of a 'deep' vocabulary learning strategies and improvement of EFL learners' vocabulary retention? RQ2: Is there any significant difference between vocabulary retention of female and male Iranian EFL learners after their being exposed to the explicit instruction of 'deep' vocabulary learning strategies?
Participants	32 EFL learners (Iranian L1) (gender mix not stated)
Setting	An EFL institute in Iran (two intact classes)
Methodology	Experimental (pre- and post-test)
Gender as a variable	Difference in vocabulary retention between genders
Method of collecting information on gender	Not stated
Results (gender)	Females in the experimental group improved their vocabulary retention ($p = 0.05$)

Study number	S7
Authors and year of publication	Garcia de Blakeley et al. (2017)
Title	Second language anxiety among Latino American immigrants in Australia
Research question(s)	RQ1: Does SLA [second language anxiety] exist among adult immigrants? RQ2: How severe is SLA among immigrants? RQ3: To what extent is SLA among adult immigrants related to demographic variables such as age, gender, and level of education? RQ4: What is the relationship between SLA and self-rated L2 competence? RQ5: To what extent is SLA among adult immigrants related to personality variables such as extroversion/ introversion and emotional stability/ neuroticism?
Participants	190 Latin American immigrants to Australia (Spanish L1) (72 males, 147 females)
Setting	Australia (general population of immigrants)
Methodology	Questionnaires
Gender as a variable	Relationship with SLA

Method of collecting information on gender	A questionnaire Detail not provided
Results (gender)	No significant differences in total SLA score between genders (p = 0.241)

Study number	S8
Authors and year of publication	Ghasemi et al. (2023)
Title	Individual differences in teacher hopelessness: examining the significance of personal and professional factors
Research question(s)	RQ1: What is the prevalence of experienced hopelessness in teachers working in the private sector in Iran? RQ2: Is there any significant difference in the levels of teachers' hopelessness in terms of their gender? RQ3: Is there any significant difference in the levels of teachers' hopelessness across their educational levels? RQ4: Is there any significant difference in the levels of teachers' hopelessness across their ages? RQ5: Is there any significant difference in the levels of teachers' hopelessness across their professional experience? RQ6: Is there any significant difference in the levels of teachers' hopelessness across the levels of education they serve?
Participants	297 EFL teachers (200 males, 97 females)
Setting	Private language schools in Iran
Methodology	Questionnaires
Gender as a variable	Difference in teacher hopelessness between genders
Method of collecting information on gender	A questionnaire Detail not provided
Results (gender)	No significant difference between levels of hopelessness and gender (p > 0.05)

Study number	S9
Authors and year of publication	Gordon (2004)
Title	"I'm tired. You clean and cook." Shifting gender identities and second language socialization
Research question(s)	Not stated

	"This article investigates the interplay between gender identity shifts and second language socialization, documenting the process by which working-class Lao women and men redefine gender identities in the United States." p. 437
Participants	Focused on 2 participants (2 females)
Setting	Laos, and Lao-American community in the US
Methodology	Interviews + participant observations + family visits + studies of Laos literacy + other qualitative data from the community
Gender as a variable	How women's gender identity shifted as a result of moving to the US/ language acquisition
Method of collecting information on gender	Not stated
Results (gender)	Women became more willing to ask for help, and their socialisation opportunities increased as a result of domestic language events

Study number	S10
Authors and year of publication	İlya (2023)
Title	Factors affecting Turkish EFL teachers' level of burnout: A quantitative study
Research question(s)	RQ1: Do English language teachers in Türkiye feel burned out as revealed by their scores on the three subscales of MBI-Educators Inventory; emotional exhaustion, depersonalization, and personal accomplishment? RQ2: Are there any effects of gender, age, length of experience, and type of school (state or private institutions) on the perceived burnout of English language teachers in Türkiye?
Participants	132 English language teachers (28 males, 104 females)
Setting	State and private schools in Turkey
Methodology	A survey
Gender as a variable	Does gender have an effect on teachers' perceived burnout
Method of collecting information on gender	Detail not provided ("demographic information was demanded from the participants" p. 146)
Results (gender)	No interaction between gender and level of teacher burnout

Study number	S11
Authors and year of publication	Inada (2021)
Title	Are there Gender differences in anxiety and any other factors among English as a foreign language (EFL) college students in Japan?
Research question(s)	Are there any gender differences in anxiety or any other factors among Japanese EFL college students?
Participants	252 students (63 males, 189 females)
Setting	A private university in Japan
Methodology	A questionnaire (five-point Likert scale)
Gender as a variable	Gender is the main variable
Method of collecting information on gender	A questionnaire Gender item was dichotomous (male vs female)
Results (gender)	Females had higher levels of anxiety than males (t (250) = -2.893, p = 0.004) Females spent more years studying English than males (t (250) = -2.288, p = 0.023)

Study number	S12
Authors and year of publication	Ju (2009)
Title	In-service adult learners' English learning strategies in Taiwan
Research question(s)	RQ1: How frequent do the adult students in this study use language learning strategies? RQ2: What are the most and the least used language learning strategies? RQ3: What is the relationship between language learning strategies used and their English proficiency level? RQ4: Does the explicit grammar instruction make any improvements in mastery the English language?
Participants	Phase 1: 184 students Phase 2: 79 students (gender mix not stated)
Setting	In-service university courses for adults in Taiwan
Methodology	Phase 1: Strategy Inventory for Language Learning and the personal background Information questionnaire Phase 2: An intervention with pre- and post-tests
Gender as a variable	Effect of gender on language learning strategy use

Method of collecting information on gender	A personal background information questionnaire Detail not provided
Results (gender)	No statistically significant effect gender on language learning strategy use

Study number	S13
Authors and year of publication	Kentmen et al. (2023)
Title	Assessing tertiary Turkish EFL learners' pragmatic competence regarding speech acts and conversational implicatures
Research question(s)	RQ1: Is there a difference between the test results of the EFL learners' production of speech acts and comprehension of implicature? RQ2: Does the comprehension of implicature differ in terms of the participants' proficiency levels? RQ3: Does the performance of speech acts vary according to students' levels of English proficiency? RQ4: Is there a difference between the female and male participants' pragmatic competence?
Participants	54 Turkish EFL learners (19 males, 35 females)
Setting	A university in Cyprus
Methodology	A discourse completion test, and on a multiple-choice discourse test
Gender as a variable	Effect of gender on a discourse completion test, and on a multiple-choice discourse test
Method of collecting information on gender	Not stated
Results (gender)	"Although there is a difference between the male and female participants in favor of male students, this difference is not considered to be statistically significant at the $p > 0.05$ alpha level with regard to the p-level of 0.8979 ($p = 0.8979$)." p. 12

Study number	S14
Authors and year of publication	Khaneshan & Bonyadi (2016)
Title	The investigation of compliment response patterns across gender and age among advanced EFL Learners

Research question(s)	RQ1: How do female and male advanced EFL learners differ regarding compliment response speech act? RQ2: How do adult and teenage advanced EFL learners differ regarding compliment response speech act?
Participants	100 Iranian EFL learners (50 males, 50 females)
Setting	An English institute in Iran
Methodology	First Certificate in English, a discourse completion task, and a compliment response framework
Gender as a variable	Effect of gender on compliment response speech act
Method of collecting information on gender	Participants were chosen by the researcher based on their gender Participants "were required to mention their gender and age in the specified blanks in the questionnaire as well" p. 762
Results (gender)	Slight, but no noticeable differences in the type of compliment strategies used Different terminology used by people of different gender to respond to compliments

Study number	S15
Authors and year of publication	Kutuk et al. (2022)
Title	Relations between gender stereotyping and foreign language attainment: The mediating role of language learners' anxiety and self-efficacy
Research question(s)	RQ1: Do Turkish EFL learners' gender stereotypes about EFL learning relate to their language attainment through self-efficacy and anxiety? RQ2: Do Turkish EFL learners' perceptions of their teachers' gender stereotypes about EFL learning relate to their language attainment through self-efficacy and anxiety?
Participants	701 EFL learners (49.4% males, 50.6% females)
Setting	Three universities in Istanbul, Turkey
Methodology	Questionnaire of Gender Stereotypes in Language Learning + Multidimensional Language Class Anxiety Scale + Questionnaire of Self-efficacy in Learning a Foreign Language + English examination scores + a demographics questionnaire
Gender as a variable	Effects of attainment in different gender groups on gender stereotypes, and teacher gender stereotypes

Method of collecting information on gender	A demographics questionnaire Detail not provided
Results (gender)	"Women who believed that EFL learning is a female domain reported higher self-efficacy, and men with the same stereotype reported lower self-efficacy." p. 223 "Paths from learner perceptions of teacher stereotypes to anxiety and self-efficacy were statistically significant, but only for women." p. 223 ($p < 0.001$)

Study number	S16
Authors and year of publication	Liang & Kelsen (2017)
Title	Gender differences in university EFL students' language proficiency corresponding to self-rated attention, hyperactivity and impulsivity
Research question(s)	RQ1: Do self-reported attention deficit, hyperactivity and impulsivity symptoms affect EFL test scores? RQ2: Is there a gender difference in self-reported attention deficit, hyperactivity and impulsivity symptoms and EFL test performance? RQ3: Do the self-reported attention deficit, hyperactivity and impulsivity symptoms affect listening, reading and overall scores differently?
Participants	229 EFL college students (89 males, 140 females)
Setting	A university in northern Taiwan
Methodology	An ADHD self-reported scale questionnaire, and the Soochow University English Proficiency Test
Gender as a variable	Effect of gender on self-reported attention deficit, hyperactivity and impulsivity symptoms, and EFL test performance
Method of collecting information on gender	Not stated
Results (gender)	Some significant interaction between the variables was found ("male students who were *likely ADHD* had English language performance below females who were both *likely ADHD* and *unlikely ADHD*" p. 68)

Study number	S17

Authors and year of publication	Lincoln & Rademacher (2006)
Title	Learning styles of ESL students in community colleges
Research question(s)	Do learning styles differ by English proficiency level, country of origin, or gender?
Participants	69 ESL students (mostly from Latin America, and Asian countries) (33 males, 36 females)
Setting	ESL courses in different locations in north-western Arkansas
Methodology	VARK learning style questionnaire (administered in different languages)
Gender as a variable	Effect of gender on learning styles
Method of collecting information on gender	Not stated
Results (gender)	Some significant effects of gender on learning styles ("Females choose the aural learning style more often than males, while males choose note taking more often than females." p. 497)

Study number	S18
Authors and year of publication	Minavandchal & Salimi (2021)
Title	Predicting EFL learners' susceptibility to various disfluency types based on gender and age
Research question(s)	RQ1: Does Iranian English learners' gender predict the production rate of each disfluency type in their speech? RQ2: Does Iranian English learners' age predict the production rate of each disfluency type in their speech? RQ3: Which disfluency type Iranian English learners are more likely to produce in their speech based on their gender and age?
Participants	40 advanced speakers of English as a foreign language (20 males, 20 females)
Setting	Iran
Methodology	Linguaskill as a proficiency measure; online, semi-structured interviews
Gender as a variable	Gender as a predictor of disfluency types
Method of collecting information on gender	Not stated Non-random sampling: participants were chosen from a pool of volunteers

Results (gender)	Gender is not a statistically significant (p = 0.049) "predictor of the production rate of filled pauses in speech" p. 186 Gender is a statistically significant (p = 0.001) "predictor of the production rate of hesitations in speech" p. 187 Male participants are more likely to produce insertions

Study number	S19
Authors and year of publication	Muñoz (2020)
Title	Boys like games and girls like movies: Age and gender differences in out-of-school contact with English
Research question(s)	RQ1: How much contact, and through what type of activities, does a sample of EFL learners from Catalonia (Spain) have outside the classroom? RQ2: Are there age and gender differences in the choice and frequency of out-of- school contact with English? RQ3: Is there an association between out-of-school contact and (self-reported) English classroom grades?
Participants	3,048 EFL learners (1,261 males, 1,787 females)
Setting	58 educational centres in Catalonia, Spain
Methodology	A questionnaire
Gender as a variable	The effect of gender on the choice and frequency on out-of-school contact with English
Method of collecting information on gender	Not stated (One of the questionnaires asked for "biographical information")
Results (gender)	Females talk in English face-to-face, and read more often, males play video games more often + other significant findings where gender, and age were combined

Study number	S20
Authors and year of publication	Natividad & Batang (2018)
Title	Students' perceptual learning styles and attitudes toward communicative language teaching
Research question(s)	RQ1: whether or not learners' ages, sexes, courses and ethnicities reveal something about their

	perceptual learning styles and their attitudes toward communicative language teaching; RQ2: whether or not there are significance differences or relationships among these learning styles and attitudes across ages, sexes, courses and ethnicities; RQ3: whether or not these profiles reveal any significant relationship across different learning styles such as visual, tactile and auditory, among others.
Participants	163 ESL students (more males than females, but data not provided)
Setting	A state university in the Philippines
Methodology	A Perceptual Learning Styles Preference Questionnaire Inventory + an inventory on attitudes towards communicative language teaching
Gender as a variable	Effects of gender on perceptual learning styles, and attitudes towards communicative language teaching *Using the term *sex* and *gender* interchangeably
Method of collecting information on gender	Not stated
Results (gender)	No differences between genders with regards to its effect on learning styles No significant differences between genders with regards to its effect on attitudes towards communicative language teaching (No *p*-values were reported)

Study number	S21
Authors and year of publication	Nikoopour et al. (2017)
Title	The impact of explicit integrated strategies instruction on IELTS applicants' listening comprehension
Research question(s)	Not stated “The present study attempted to explore the impact of the explicit instruction of cognitive and memory strategies on IELTS candidates' listening comprehension” p. 774
Participants	88 IELTS candidates (gender mix not stated)
Setting	A language institute in Tehran, Iran
Methodology	A proficiency test + a pre-test, intervention, post-test

Gender as a variable	"A moderator variable" – does it make a difference in listening comprehension
Method of collecting information on gender	"gender was included as a moderator variable; thus, both male and female candidates were selected" p. 776 – not stated how this was determined
Results (gender)	No statistically significant finding

Study number	S22
Authors and year of publication	Pichette et al. (2019)
Title	Detection of gender-biased items in the Peabody Picture Vocabulary Test
Research question(s)	Not stated "The primary goal of this study is to identify gender bias on a sample language test (the PPVT-IV)" p. 7
Participants	443 ESL learners (133 males, 310 females)
Setting	A university
Methodology	The Peabody Picture Vocabulary Test
Gender as a variable	The effects of gender on the PPVT-IV scores
Method of collecting information on gender	Not stated
Results (gender)	Males scored significantly higher than females on the PPVT-IV ($t = 5.4$; $p < 0.001$)

Study number	S23
Authors and year of publication	Reigel (2008)
Title	Positive feedback in pairwork and its association with ESI course level promotion
Research question(s)	RQ1: How might the total rate of positive feedback received have affected these outcomes for the students? RQ2: Is it possible that students who were promoted generally enjoyed higher frequencies of feedback? RQ3: If so, under what circumstances are students receiving more feedback?
Participants	44 adult ESL learners (gender mix not stated)
Setting	Portland State University Laboratory School
Methodology	Analysis of audio recordings from the lessons
Gender as a variable	Effect of gender on positive feedback rate, and level promotion

Method of collecting information on gender	Not stated (Researchers were provided with the students' names)
Results (gender)	Gender did not have any significant effects on positive feedback rate, nor on the likelihood to be promoted up a level

Study number	S24
Authors and year of publication	Renani et al. (2019)
Title	Effect of awareness of teacher education philosophy on EFL teachers' professional knowledge: A postmethod perspectivisation
Research question(s)	RQ1: Does EFL teachers' awareness of TEP [teacher education philosophy] affect the PK [professional knowledge] of male and female teachers? RQ2: Does EFL teachers' awareness of TEP affect their PK at two levels of academic qualification? RQ3: Does EFL teachers' awareness of TEP affect their PK at three levels of teaching experience? RQ4: Does EFL teachers' awareness of TEP affect their PK in three age ranges?
Participants	60 EFL teachers (gender mix not stated)
Setting	Four institutes in Iran
Methodology	An experimental study with a control group + Philosophy of Adult Education Inventory + a PK questionnaire + observations
Gender as a variable	The effect of participants' TEP on the PK were compared within gender groups (female control vs female experimental group; male control vs male experimental group)
Method of collecting information on gender	Not stated
Results (gender)	No comparison of genders carried out

Study number	S25
Authors and year of publication	Rezvani & Tavakoli (2013)
Title	Investigating Iranian test-takers' cognitive and metacognitive strategy use: IELTS reading section in focus
Research question(s)	RQ1: Is there any relationship between Iranian test-takers' use of cognitive and metacognitive strategies

	and their EFL reading test performance on the reading section of the IELTS test? RQ2: Is there a significant difference between male and female Iranian test-takers in terms of their use of cognitive and metacognitive strategies?
Participants	52 advanced EFL learners who recently completed an IELTS preparation course (30 males, 30 females at the start)
Setting	Iran
Methodology	IELTS Academic + cognitive, and metacognitive strategy questionnaire
Gender as a variable	Effect of gender on cognitive, and metacognitive strategy use
Method of collecting information on gender	Not stated
Results (gender)	No significant difference in strategy use between male and female participants

Study number	S26
Authors and year of publication	Sadeghi & Sharifi (2013)
Title	The effect of post-teaching activity type on vocabulary learning of elementary EFL learners
Research question(s)	Not stated “This study set out to investigate the effect of four post-teaching activities, namely game, narrative writing, role-play, and speaking tasks on vocabulary gain of elementary Iranian EFL learners across gender.” p. 65
Participants	111 elementary EFL learners (47 males, 64 females)
Setting	A language school in Iran
Methodology	Cambridge Key English Test + a vocabulary pre-test, an intervention, a vocabulary post-test
Gender as a variable	Relationship between gender and post-teaching activity type on vocabulary gains Participants were split into groups based on gender
Method of collecting information on gender	Not stated
Results (gender)	Gender had a significant impact with female participants outperforming males ones ($p = 0.000$); gender had a significant impact on vocabulary learning ($p = 0.006$)

Study number	S27
Authors and year of publication	Shang (2016)
Title	Exploring demographic and motivational factors associated with hypertext reading by English as a foreign language (EFL) students
Research question(s)	RQ1: Students' attitudes towards hypertext learning experiences (i.e. usefulness, ease of use, and attitude of future use). RQ2: Differences in gender, age, and proficiency level on hypertext learning experience. RQ3: Relationships existing among different genders, ages, proficiency levels, and hypertext learning experiences on English reading comprehension.
Participants	23 EFL students (9 males, 14 females)
Setting	A private university in Taiwan
Methodology	A General English Proficiency Test (the reading component) + hypertext exposure with a content knowledge test + a perceptions questionnaire
Gender as a variable	Effects of gender on hypertext learning experience
Method of collecting information on gender	A demographic questionnaire Detail not provided
Results (gender)	"females perceived to be able to reduce more reading time in hypertext use … than males" (p = 0.044), and females found it easier to do the task (p = 0.019) p. 565 "The overall effect size of hypertext learning experience was [d =] 0.64, a large difference favouring females" p. 566

Study number	S28
Authors and year of publication	Soltanian & Sadeghi (2021)
Title	Thin-slice judgments of English language teacher success in instruction: The effects of learners' gender, age, and language proficiency
Research question(s)	RQ1: To what extent do EFL learners' impressions of their language teachers differ at the beginning and at the end of the semester? RQ2: Are there any significant differences between EFL learners' degree of FI change and their

	demographic characteristics such as gender, age, and English proficiency level?
Participants	679 EFL learners (297 males, 382 females)
Setting	Five language institutes, and four universities in Iran
Methodology	Characteristics of Successful Teachers Questionnaire administered at two different time points
Gender as a variable	Effects of gender on first impressions of language teachers
Method of collecting information on gender	Not stated
Results (gender)	"There is a significant difference between females and males in the extent of FI (first impressions) change" p. 8 ($p < 0.05$)

Study number	S29
Authors and year of publication	Tabatabaei et al. (2020)
Title	A psychoanalytic health survey on the Iranian EFL learners' emotional intelligence and their language learning approach
Research question(s)	RQ1: Is there any significant relationship between Iranian EFL learners' emotional intelligence and their vocabulary strategy use? RQ2: Do Iranian EFL learners with higher emotional intelligence perform better in vocabulary strategy use? RQ3: Is there any difference between Iranian EFL male and female learners in vocabulary strategy use?
Participants	60 medicine students/ EFL learners (30 males, 30 females)
Setting	Yasuj Univeristy of Medical Sciences in Iran
Methodology	Bar-On Emotional Quotient Inventory + Strategies for Vocabulary Learning questionnaire
Gender as a variable	Effects of gender on vocabulary strategy use
Method of collecting information on gender	Not stated (participants were selected)
Results (gender)	A positive correlation between gender, and vocabulary strategy use ($p = 0.044$); males scores statistically significantly higher than females ($p = 0.000$)

Study number	S30
Authors and year of publication	Tavakoli & Davoudi (2017)
Title	Willingness to communicate rally: The case of Iranian EFL learners
Research question(s)	RQ1: Does the questionnaire on willingness to communicate (WTC) orally have an acceptable index of reliability and validity among Iranian EFL learners? RQ2: Which factor (age, gender, or interlocutor) has the greatest influence on Iranian EFL students' WTC orally?
Participants	117 EFL learners (40 males, 77 females)
Setting	A language school in Iran
Methodology	A questionnaire on oral WTC
Gender as a variable	Effects of gender on students' oral WTC
Method of collecting information on gender	Not stated
Results (gender)	No significant effect of gender found

Study number	S31
Authors and year of publication	Tenney et al. (2020)
Title	Assessment of English language performance scores and academic performance in an English-based curriculum for pharmacy students with English as a second language
Research question(s)	Not stated "The primary objective was to determine if there is a relationship between English language performance and graduating grade point average (GPA) in pharmacy students with English as a second language (ESL)." P. 423
Participants	113 pharmacy students, L1 Chinese (51% males, 49% females); ESL speakers
Setting	A public university in Hong Kong
Methodology	Comparing pre-admission results (including English), and academic performance
Gender as a variable	Exploring whether the effects hold true for both genders
Method of collecting information on gender	"Demographic information was collected including gender" p. 424 Detail not provided

Results (gender)	Pre-admission English scores were a stronger predictor of academic success for women (p = 0.009) than men (p = 0.053)

Study number	S32
Authors and year of publication	Tercanlioglu (2004)
Title	Exploring gender effect on adult foreign language learning strategies
Research question(s)	RQ1: What is the mean level of students' FL learning strategy use? RQ2: Are the scales of "Strategy Inventory for Language Learning" correlated with each other? RQ3: Is there a statistically significant gender difference in students' FL learning strategy use?
Participants	184 Year 4 students, pre-service EFL teachers (44 males, 140 females)
Setting	Atatürk University in Turkey
Methodology	Strategy Inventory for Language Learning
Gender as a variable	Effects of gender on language learning strategy use
Method of collecting information on gender	Not stated
Results (gender)	Males used some strategies more than females, and vice versa (significance ranging from $p < 0.05$ to $p < 0.0001$)

Study number	S33
Authors and year of publication	Tian et al. (2022)
Title	Chinese EFL learners' attitudes towards smartphone-based reading
Research question(s)	Not stated “The present research conducted a questionnaire survey to explore EFL learners' perceptions and beliefs by integrating the reading attitude model with the technology acceptance model UTAUT2.” p. 1838
Participants	192 EFL learners (62 males, 130 females)
Setting	A local university in China
Methodology	An attitudes questionnaire
Gender as a variable	Effects of gender on smartphone-based reading attitudes

Method of collecting information on gender	A questionnaire Detail not provided
Results (gender)	Males were less influenced by famous online figures than females (p = 0.036), and they found annotating on smartphones easier than females (p = 0.024)

Study number	S34
Authors and year of publication	Weger (2013)
Title	International students' attitudes toward L2-English classroom activities and language skills in the USA
Research question(s)	RQ1: What underlying constructs are present in the responses of international students studying L2-English abroad to a questionnaire regarding their preferences for skill-focused classroom activities? RQ2: Are there variations in the classroom activity preferences across these learners on the basis of gender, age, or course level? RQ3: Which skill-focused classroom activities do these learners prefer?
Participants	131 ESL learners (55 males, 76 females)
Setting	An intensive ESL programme at a private university in Washington, DC, the US
Methodology	A questionnaire on classroom activity preferences
Gender as a variable	Effects of gender on classroom activity preferences
Method of collecting information on gender	A biographical data form was used, not explicitly stated whether this included information on gender
Results (gender)	No significant effect of gender on classroom activity preferences

Study number	S35
Authors and year of publication	Xue (2021)
Title	The developmental trajectory of biliteracy for Chinese–English adult EFL learners: a longitudinal study
Research question(s)	RQ1: What was the developmental trajectory of biliteracy and related language/ cognitive skills for Chinese EFL learners in a period of 9-month EFL learning?

	RQ2: What was the relationship between L1 literacy skills, L2 literacy and related language/cognitive skills with the increase of L2 proficiency?
Participants	139 EFL college students (19 males, 120 females)
Setting	A university in China
Methodology	Literacy tests + literacy-related language and cognitive skills tests
Gender as a variable	Relationship between gender, biliteracy, and related language/ cognitive skills
Method of collecting information on gender	Not stated
Results (gender)	No significant effect of gender (p = 0.64 at pre-test; p = 0.61 at post-test)

Study number	S36
Authors and year of publication	Zhang & Ardasheva (2019)
Title	Sources of college EFL learners' self-efficacy in the English public speaking domain
Research question(s)	RQ1: What are, overall, the relative contributions of sources of self-efficacy to EPS self-efficacy among Chinese college EFL learners? RQ2: Do the relative contributions of the four sources vary by EPS course experience? RQ3: Do the relative contributions of the four sources vary by gender? RQ4: Do the relative contributions of the four sources vary by academic major?
Participants	263 EFL college students (60 males, 203 females)
Setting	Six Chinese universities
Methodology	Two self-efficacy scales + a demographic questionnaire
Gender as a variable	Effects of gender on the relative contributions of sources of self-efficacy
Method of collecting information on gender	A questionnaire Gender item was dichotomous (male vs female)
Results (gender)	Different sources of self-efficacy explained variance in English public speaking self-efficacy for males and females EME accounted for 8% variance in M (p < 0.001), 20% in F (p < 0.05)

	VE accounted for 3% variance in F ($p < 0.01$), marginally significant for M ($p < 0.07$) VP accounted for 5% variance in F ($p < 0.001$), not significant for M No significant results for PAS

Appendix 3: RQ2 methodology by country

Country	Questionnaire: three choices	Questionnaire: dichotomous choice	Blanks & chosen based on gender	Questionnaire (detail not provided)	Not stated	Not stated but demographic info requested	Not stated & chosen based on gender
Iran (n = 12)	-	-	1	2	6	-	3
Taiwan (n = 4)	-	-	-	2	2	-	-
The US (n = 4)	-	-	-	-	3	1	-
China (n = 3)	-	1	-	1	1	-	-
Spain (n = 3)	1	-	-	1	-	1	-
Turkey (n = 3)	-	-	-	1	1	1	-
Australia (n = 1)	-	-	-	1	-	-	-
Cyprus (n = 1)	-	-	-	-	1	-	-
Germany (n = 1)	-	1	-	-	-	-	-
Hong Kong (n = 1)	-	-	-	-	-	1	-
Japan (n = 1)	-	1	-	-	-	-	-
The Philippines (n = 1)	-	-	-	-	1	-	-
Not stated (n = 1)	-	-	-	-	1	-	-
Total (n = 36)	1	3	1	8	16	4	3

Appendix 4: Overview of qualitative text interpretation for RQ3

Study	Context	Focus	Methodology	Significant findings	In line with the definition of gender
S2	Spain, university	Uptake of Language Massive Open Online Courses (LMOOC)	A questionnaire	Female learners are more likely to take the courses than males	✅ Enrolling onto a course can be viewed as something one does, and it can be socio-culturally driven. This can be context-specific. *The method does not include an interaction between participants.
S4	Germany, general population	Acquisition of L2 vernacular variation	Semi-structured interviews	Females use innovative variants more readily (factor weight FW = 0.64; $p < 0.0349$)	✅ The way we speak and adjust our language is something one does in an interaction to conform to socio-cultural expectations. This can be context-specific.
S6	Iran, language school	Vocabulary retention/ vocabulary learning strategies	Experimental (pre-test, intervention, post-test)	Females in the experimental group improved their vocabulary retention ($p = 0.05$)	❌ This is not performed in an interaction, the strategies aren't used under socio-cultural influence.
S9	The US, general population	Gender identity shifts	Interviews + participant observations + family visits + studies of Laos literacy + other qualitative data	Women became more willing to ask for help, and their socialisation opportunities, and economic independence increased as a result of	✅ The observed gender identity shifts are a result of what happened in interactions, and they are heavily influenced by socio-cultural expectations. This is context-specific.

Study	Context	Focus	Methodology	Significant findings	In line with the definition of gender
			from the community	domestic language events	
S11	Japan, university	Foreign language anxiety	A questionnaire (self-assessment)	Females had higher levels of anxiety than males (t (250) = -2.893, *p* = 0.004) Females spent more years studying English than males (t (250) = -2.288, *p* = 0.023)	✅ Anxiety in foreign language use/ classroom is something that one does in an interaction, and hence it could be socio-culturally influenced. This can be context-specific. *The method does not include an interaction between participants.
S14	Iran, language school	Compliment response patterns	A discourse compliment task	Males and females used different language to enact compliment response	✅ Compliment response is what one does in an interaction, which is both socio-culturally influenced, and context-specific. *The method does not include an interaction between participants, but draws on it.
S15	Turkey, universities	Gender stereotypes, and language attainment	Questionnaire of Gender Stereotypes in Language Learning + Multidimensional Language Class Anxiety Scale +	Women's beliefs that EFL learning is a stereotypically female domain led to higher self-efficacy, and to higher attainment. Perception of teacher stereotypes, self-efficacy, and anxiety	✅ Perceptions of gender stereotypes are socio-culturally constructed, and they are context-dependent. It is something one creates based on an interaction. *The method does not include an interaction between participants, but reports on it.

Study	Context	Focus	Methodology	Significant findings	In line with the definition of gender
			Questionnaire of Self-efficacy in Learning a Foreign Language + English examination scores	were statistically significant for women ($p < 0.001$).	
S16	Taiwan, university	Attention, hyperactivity, and impulsivity	A questionnaire (self-reported ADHD) + a proficiency test	Males with likely ADHD performed worse than both likely, and unlikely ADHD females; ADHD results in lower test performance in males, but not in females	❌ ADHD is a psychiatric diagnosis, and despite its effect on behaviour, it is not something that one does in an interaction, and it is not influenced by socio-cultural expectations. This is unlikely to be context-specific. *The method does not include an interaction between participants.
S17	The US, community colleges	Learning styles	VARK learning style questionnaire	Males stated a preference for a kinaesthetic style, females for aural learning style. Learning tyles differ for Hispanic and Asian males.	❌⚠️ Learning style preferences is not something one does in an interaction, they are cognitive, and unlikely to be influenced by socio-cultural expectations. However, the self-reported nature of the questionnaire might result in gender stereotypical answers. The results also suggest that the results are context-dependent.

Study	Context	Focus	Methodology	Significant findings	In line with the definition of gender
S18	Iran, a pool of volunteers	Disfluency types	Semi-structured interviews	Males are more likely to produce insertions than deletions, females hesitations (p = 0.001)	✅ Disfluency is something that one does in an interaction, and it is context-dependent – participants' actions are guided by socio-cultural expectations.
S19	Catalonia, Spain, educational centres	Out-of-school contact with English	A questionnaire	Females talk in English face-to-face, and listen to music more often than males. Males play video games more often. Adolescent females read, and watch movies/ series more often.	✅ Out-of-school contact with English is something one does – not necessarily always in an interaction, but it is likely to be heavily influenced by socio-cultural expectations, and context-dependent. *The method does not include an interaction between participants.
S22	Not stated, university	Item detection in the Peabody Picture Vocabulary Test IV (The full test done in groups)	Peabody Picture Vocabulary Test IV (The full test done in groups)	Males scored significantly higher than females (t = 5.4; p < 0.001), 3% of items showed a difference in detection by gender	✅❌⚠️ Performance on a vocabulary test is not something one does in an interaction, and it is not socio-culturally influenced. However, contrary to how the PPVT is normally administered, the researchers continued with the test beyond the point where the participants couldn't answer the questions anymore. This resulted in them guessing the answers, which can be interpreted as socio-culturally influenced behaviour. Context can have influence on the familiarity of the test items.

Study	Context	Focus	Methodology	Significant findings	In line with the definition of gender
S26	Iran, language school	Post-teaching activity types, and vocabulary learning	Experimental (pre-test, intervention, post-test)	Gender had a significant impact on vocabulary learning: female participants outperformed males (p = 0.000)	❌ Vocabulary acquisition is more cognitive than behavioural. Whilst some activity types included interaction, and the context (single sex experimental groups) could have had an impact on the learners' behaviour because of socio-cultural expectations, the post-test was delayed, so the interaction in itself cannot be isolated as a correlative factor.
S27	Taiwan, university	Hypertext learning experience	Hypertext reading task + a perceptions questionnaire	Females felt that they were able to reduce reading time in hypertext use more than males (p = 0.044), and found it easier to do the task than males (p = 0.019) Females scored higher on hypertext learning experience than males (d = 0.64)	❌ Reading is not done in an interaction, and despite the existence of gender stereotypes, reading is not a socio-culturally conditioned behaviour. This could be context-dependent, with context having an impact on the availability of reading material, but this contextual factor would not be related to gender.
S28	Iran, language schools + universities	First impressions of language teachers	Characteristics of Successful Teachers Questionnaire administered at	Males and females change their first impressions of English language teachers to a different extent (p < 0.05)	✅ Thin-slice judgements are inherently what one does in a social interaction, and they are influenced by socio-cultural norms/ expectations. These would vary between different contexts.

Study	Context	Focus	Methodology	Significant findings	In line with the definition of gender
			two different time points		*The method does not include an interaction between participants.
S29	Iran, university	Emotional intelligence and vocabulary strategy use	Bar-On Emotional Quotient Inventory + Strategies for Vocabulary Learning questionnaire	A positive correlation between gender, and vocabulary strategy use (p = 0.044); males were more correlated with vocabulary strategy use than females (p = 0.000)	❌ Vocabulary learning strategies are not something that one does in an interaction, so they are not likely to be susceptible to socio-cultural expectations. They are unlikely to be context-dependent.
S31	Hong Kong, university	Academic performance	Pre-admission tests + grade point average (GPA)	Pre-admission English scores were a stronger predictor of academic success for women (p = 0.009) than men (p = 0.053)	❌ ⚠ Correlations between pre-admission results, and academic attainment are not something one does in an interaction. The longitudinal design means that socio-cultural expectations might have, however, had an impact on the participants' study behaviour. This can be context-dependent. *The method does not include an interaction between participants.
S32	Turkey, university	Language learning strategies	Strategy Inventory for Language Learning	Females reported higher use of 15 out of 50 individual strategies (from p<0.05 to p<0.0001). Males reported higher scored	⚠ Language learning strategies are not something that one does in an interaction, so they are not likely to be susceptible to socio-cultural expectations. However, some, such as "Managing your emotions" could be

Study	Context	Focus	Methodology	Significant findings	In line with the definition of gender
				in “Using all your mental processes” and “Organising and evaluating your learning ($p < 0.001$ and $p < 0.05$ respectively).	influenced by socio-cultural expectations which could be context-dependent. *The method does not include an interaction between participants.
S33	China, university	Attitudes towards smartphone-based learning	A questionnaire	Males were less influenced by famous online figures than females ($p = 0.036$), and they found annotating on smartphones easier than females ($p = 0.024$).	✅ Attitudes are susceptible to the influence of socio-cultural expectations, and they are borne out of both interactions, and context. *The method does not include an interaction between participants.
S36	China, universities	Sources of self-efficacy	Two self-efficacy scales	Different sources of self-efficacy explained variance in English public speaking self-efficacy for males and females EME accounted for 8% variance in M ($p < 0.001$), 20% in F ($p < 0.05$) VE accounted for 3% variance in F ($p < 0.01$),	✅❌⚠️ Two thirds of the results can be interpreted in line with the definition of gender: VE (vicarious experience) and VP (verbal persuasion) are two sources of self-efficacy which are performed/ gained from an interaction, they are susceptible to socio-cultural expectations, and can be context-dependent, so they would be in line with the definition of gender. EME (enactive mastery experience) is not gained from an interaction, and it is not context-dependent, so this would not be in line with the definition of gender.

Study	Context	Focus	Methodology	Significant findings	In line with the definition of gender
				marginally significant for M ($p < 0.07$) VP accounted for 5% variance in F ($p < 0.001$), not significant for M No significant results for PAS.	*The method does not include an interaction between participants.