Gender Differences in L2 Motivation among Secondary School Students in Hong Kong:

A Meta-Analysis of Studies on Learning English as a Second Language

Pui Yu Ivy Chan

MSc Applied Linguistics and Second Language Acquisition, 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]:	
Date:	6 Aug 2024

Gender Differences in L2 Motivation among Secondary School Students in Hong Kong: A Meta-Analysis of Studies on Learning English as a Second Language

Pui Yu Ivy Chan 陳沛瑜

Department of Education, University of Oxford

St. Hugh's College

Supervisor: Dr. Anna-Maria Ramezanzadeh

Dissertation submitted in part-fulfillment of the requirements for the degree of Master of Science in Applied Linguistics and Second Language Acquisition

Trinity Term, 2024

Word count: 19,655

Abstract

This meta-analysis synthesizes existing research on gender differences in L2 motivation among secondary school students in Hong Kong. A total of eight research reports, involving 24,660 language learners, were meta-analyzed. The results showed that studies on general L2 motivation found a significant, small effect size in favour of females (p=0.00, d=-0.226), indicating that females were slightly more motivated than their male counterparts in learning English. However, studies concerning specific English skills (i.e., reading motivation, writing motivation) only found a very small gender difference (p=0.04, d=-0.037), indicating that the gender effect may not be observed in the motivation to learn specific language skills. Overall, the combined effect size across studies was significant, but very small (p=0.00, d=-0.161).

This study also examined the moderating effects of four factors: (i) age, (ii) medium of instruction, (iii) school banding, and (iv) L2 proficiency. The results suggested a significant age effect on the gender difference (t(1)=-3.178, B=-0.149, F=10.100, p=0.025): the older the participants, the wider the gender gap in L2 motivation. The gender gap was found to emerge in Secondary 3, but the difference became insignificant again in Secondary 6. No significant effect was observed for the medium of instruction, school banding, and L2 proficiency, possibly due to missing data.

This meta-analysis provided novel insights into the criticality of adolescence in L2 motivation, adding crucial evidence to the discussion of gender divergence in language abilities during teenage years. Given the observed differentiation in Secondary 3, it is recommended that teachers devise focused motivational strategies, enhance career consultations for students, and participate in ongoing professional development to create gender-inclusive L2 classrooms.

Keywords: motivation, second language, English, gender, systematic review, meta-analysis, Hong Kong, secondary school, teenagers

Acknowledgements

Words cannot express my gratitude to my supervisor, Dr. Anna-Maria Ramezanzadeh, for being the best tutor one could ask for. Thank you for dedicating hours discussing, refining, and providing constructive feedback on our projects. Your expertise in learning motivation has, and will continue to inspire my growth in both teaching and scholarly pursuits.

I would be remiss in not also thanking my family. Their boundless faith and trust has enabled me to roam free in the big, vast world, pursuing my passions without reservation. Special thanks goes to my very intelligent brother, Samson, for his constant and timely statistics support. Your steadfast encouragement and witty humour has buoyed me through the arduous storms.

To my friends at CUHK, especially my archery coach, Johnson: you all are my reason and strength to carry on. I still vividly remember the night I received the offer from Oxford, reading the email as I retrieved my arrows. You were the first to bear the news, and just as you've shared my joy, you've been there in times of sorrow. Despite the time difference, you were always behind the camera when I tuned in to Zoom for archery training. Thank you for rejuvenating me during the darkest days, and for making me feel at home 6,000 miles away.

My appreciation also goes to my pals from HYS, especially Christine and Judy, who have read and provided invaluable feedback on all of my assignments and chapters. Your incredible strength, independence, and intelligence as young women were what sparked my interest in gender studies in the first place.

Lastly, to Alex, Chen, David, Felicity; my systematic review pals Eman, Jolin, Ivy; and all the friends I have made on the way: your friendship has been my greatest gift from Oxford. Together, we spent countless days locked in libraries, spontaneously presenting and debating dissertation ideas; equally, we spent (too) much time indulging in endless Smash games at the MCR, late-night conversations, and exciting trips to Barcelona and Port Meadow. Your companionship actualised what I deemed unlikely before coming to Oxford: simple friendship in an elite, competitive environment. This dissertation is for you; for your camaraderie, wisdom, and support have made this journey not only bearable, but also joyful.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

AMTB	Attitude/Motivation Test Battery
CFA	Confirmatory Factor Analysis
CS	Current Self
ESL	English as a second language
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HKDSE	Hong Kong Diploma of Education
HKSAR	Hong Kong Special Administrative Region
IE	Intended Effort
IS	Ideal second language self
L2	Second language
L2LE	Second language learning experience
L2MSS	Second Language Motivational Self System
MoI	Medium of Instruction
OTS	Ought-to self
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QATSDD	Quality Assessment Tool for Studies with Diverse Designs
QuADS	Quality Assessment with Diverse Studies
RQ	Research Question
S1	Secondary 1
SD	Standard deviation
SDT	Self-determination theory
S-EM	Socio-Educational Model
SEM	Standard error of the mean
WTC	Willingness to communicate

1. INTRODUCTION

1.1. Background and Rationale

Gender is a fundamental part of human identity, underpinning many, if not all, human interactions (West and Zimmerman 2002). With the rise of feminist narratives in academia, research concerning gender differences in education emerged in the 1960s, and has since then been viewed as a critical pathway to educational equity (Watt 2016).

The current study concerns gender differences in students' learning motivation, with a focus on English as a Second Language in Hong Kong secondary school students. Motivation has been recognized as a pivotal component of second/foreign language learning, for it not only drives the initial learning stages, but also provides impetus for sustained learning (Dörnyei 1998). Studies in the past have revealed that female learners tend to be more motivated in language arts, reporting higher self-efficacy (Eccles et al. 1993) and enjoyment (Watt 2004) in L2 learning. However, research findings are far from homogenous, with some studies showing little to no gender differences in L2 motivation (e.g., Cortés 2002; Pang 2014; Ravid 1987), and some showing significant gender differences (e.g., Lai 2007; Shaaban and Ghaith 2000; Watt 2004). This points to a need for meta-analyses to synthesize these findings, and to investigate whether existing studies, combined, have found significacnt gender differences in L2 motivation.

Research on gender differences in L2 learning has pointed to the criticality of adolescence. Scholars such as Eccles (1987) and Watt (2004) argued that during adolescence, social agents like parents, peers, media, and schools may heighten boys' and girls' interest in doing gender-appropriate activities, and their perceptions of certain subjects, such as viewing language learning to be "feminine" (Watt 2004). This hypothesis of teenage divergence is echoed by Maccoby and Jacklin's (1974) meta-analysis, which found that boys' and girls' verbal abilities start to diverge during adolescence (age 12 and beyond). This points to a critical phase in language development that is worth scholarly attention, as differences in motivation during this phase might contribute to boys' and girls' divergence in L2 proficiency, potentially contributing to gender inequality in L2 education.

Hong Kong, a bilingual Asian society, provides a unique sociolinguistic landscape for the current study. Given that English is one of the core subjects for university admission, English is essential for academic and career success, which might externally motivate students to achieve a higher L2 proficiency. Gender-wise, with its British colonial history, Hong Kong gender practices involve two distinct cultural formations, integrating the Western

and Chinese conceptions of sexual differences (Ha 2009). This unique conceptualization of gender constitutes unique gender norms and practices, which might impact one's gender socialization and hence motivation in learning "gendered" subjects such as languages.

Given that gender is a social construct, gender phenomena cannot be studied out of context, for gender roles, norms, and expectations vary across cultures and regions (Kalliyath 2015). Just as the expectations for women and girls would be very different in Hong Kong and in Canada, the gender differences in L2 motivation would likely differ by contexts. Currently, there is no existing L2 motivational meta-analysis focused on Hong Kong, which is a research gap the current study aims to fill.

The heterogeneity in gender and motivation studies, coupled with the unique gender culture in Hong Kong and the criticality of adolescence in language development, underscores the lack of literature summarizing Hong Kong adolescents' gender experiences in L2 learning. These factors provide strong impetus for this meta-analysis to analyze gender differences in L2 motivation among Hong Kong secondary school students.

1.2. Outline

This meta-analysis begins with a literature review on motivational theories and relevant studies on gender and motivation in Hong Kong (Chapter 2). The research questions and methodology are then outlined (Chapter 3), followed by results and data analysis (Chapter 4). It concludes with a discussion of the findings and pedagogical implications, as well as areas for future research (Chapter 5).

2. LITERATURE REVIEW

This chapter provides an overview of the literature related to gender differences in L2 motivation in Hong Kong. It begins with a definition of the term "motivation" (Section 2.1), followed by critical evaluations of three motivational frameworks: the Socio-Educational Model (Section 2.2), Self-Determination Theory (Section 2.3), L2 Motivational Self System (Section 2.4), and a comparison between the three (Section 2.5). The chapter then discusses the body of literature on gender differences in L2 motivation (Section 2.6), focusing on Hong Kong (Section 2.7) and adolescence (Section 2.8). The chapter concludes by summarizing the research gaps (Section 2.9).

2.1. Defining "Motivation"

Motivation is a pivotal factor in L2 learning. It initiates and sustains the often long and tedious learning process, propelling language learners to accomplish their goals even when their aptitude and learning conditions are far from ideal (Dörnyei 2005). Two meta-analyses conducted by Yousefi and Mahmoodi (2022) and Masgoret and Gardner (2003), involving 17 and 75 studies respectively, showed that L2 motivation not only had a moderate effect on learning behaviour (r=0.53), but also correlated with L2 grades (r=0.37). These analyses offer strong evidence that motivation affects actual learning behaviour, and may eventually impact one's language learning outcomes.

Despite its importance in L2 learning, motivation is often difficult to define as it encompasses a variety of cognitive, affective and behavioural processes (McEown, Noels, and Chaffee 2014). In broad terms, L2 motivation can denote the reason behind language learning behaviours, the sustentation of said behaviours, and the effort invested in learning a language (Boekaerts 1995; Dörnyei and Ushioda 2021). However, given the complexity of the variables inherent to the concept, the operationalization of "L2 motivation" has been diverse. For example, Gardner (2010b), who proposed the Socio-Educational model (S-EM), defined motivation as "a long-term commitment to the task with the associated effort, desire, and satisfaction" (p.207). On the other hand, Dörnyei (2009), who proposed the L2 Motivational Self System (L2MSS), drew upon psychological studies and defined motivation as the desire to bridge the gap between one's current self-concept and ideal self-concept in language learning.

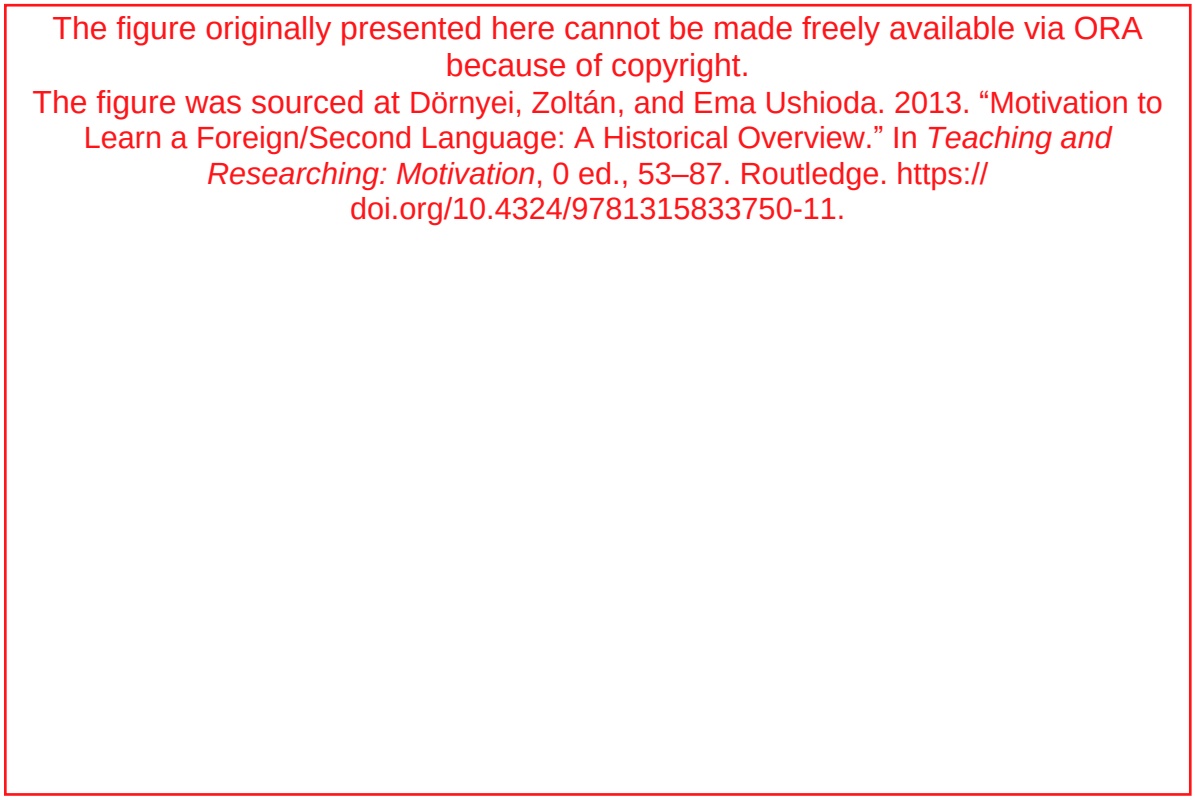
From 2005/2006 to 2011/2012, Gardner's socio-educational model (S-EM) dominated the field of motivational studies, before it was surpassed in popularity by Dörnyei's L2

Motivational Self System (L2MSS) in 2013/2014 (Boo, Dörnyei, and Ryan 2015). Comparatively, Deci and Ryan's Self-Determination Theory is less popular, but it has been consistently applied in a considerable number of studies. Sections 2.2-2.4 critically evaluate these three dominant motivation frameworks, followed by a discussion of their similarities and differences.

2.2. Gardner's Socio-Educational Model (S-EM)

2.2.1. Overview of the Socio-Educational Model

Gardner's S-EM, first proposed in 1972 with Lambert, has been highly acclaimed among L2 researchers (Dörnyei 2005). The underlying hypothesis of S-EM is simple: it posits that language learning involves not only grammatical decoding and vocabulary learning (i.e., cognitive tasks), but also the learners' willingness to identify with the members of another ethnolinguistic group and to take on very subtle aspects of their behaviour, including their distinctive style of speech and their language (Gardner and Lambert 1972).



As such, according to the S-EM, motivation involves an orientation (i.e., a purpose). The orientation can be "integrative", where the learner wishes to integrate into, interact with, or identify with members of the L2 community (Noels 2002). Alternatively, the orientation could also be "instrumental", where the learner wishes to learn the language "for pragmatic

reasons that do not involve identification with the other language community" (Gardner 2001, 8).

In addition to orientation, the learners' "attitudes toward the learning situation" also mediate their motivation for language learning (Gardner, Lalonde, and Pierson 1983). This involves their physical learning environment, such as the teacher, the course, classmates, and teaching materials. Thus, the combination of "orientation" and "attitude towards the learning situation" forms the basis of L2 motivation.

To measure these aspects of language learning, Gardner and Smyihe (1981) developed the Attitude/Motivation Test Battery (AMTB). Using 11 subscales, the instrument measures five constructs: integrativeness, attitude toward the learning situation, motivation, instrumental orientation, and language anxiety (see Table 1). Contrary to many standardized tests, Gardner (1985a) emphasizes that AMTB is not static, and that the items on AMTB should be modified to suit different research contexts. A meta-analysis by Masgoret and Gardner (2003) revealed that the subscales of AMTB are highly reliable, with Cronbach alphas ranging from 0.79 to 0.93, except for instrumental orientation, which only generated a poor reliability of 0.57. This suggests a need to reconceptualize "instrumental orientation."

Table 1. AMTB Constructs, Subscales, and Definitions (Gardner 2001)

Construct	Subscales	Definition of measure (Gardner 1985b)
Integrativeness	Integrative orientation (4 items)	The learner's self-rated importance of learning the L2 in order to interact with other L2 speakers.
	Interest in foreign languages (10 items)	General interest in studying foreign languages.
	Attitudes toward L2 community (10 items)	Attitude toward the community that speaks the L2.
Attitudes toward the learning situation	Evaluation of the L2 teacher (10 items)	General evaluation on their L2 teacher's friendliness, reliability, considerateness, goodness, pleasantness, efficiency, politeness, sincerity, dependability, and cheerfulness.
	Evaluation of the L2 course (10 items)	General evaluation on their L2 course's goodness, agreeability, pleasantness,

		satisfaction, quality, pleasantness, enjoyability, rewardingness, value, and appeal.
Motivation	Motivational intensity (10 items)	The learner's motivation to learn the L2 in terms of completed assignments, and future plans to use and study the L2.
	Desire to learn L2 (10 items)	The learner's desire to learn the L2.
	Attitudes toward learning L2 (10 items)	The learner's attitude toward learning the L2.
Instrumental orientation	Instrumental orientation (4 items)	The pragmatic or utilitarian value of learning the L2.
Language anxiety	L2 class anxiety (10 items)	Degree of discomfort while participating in the L2 class.
	L2 use anxiety (10 items)	Degree of discomfort while using the L2.

2.2.2. Support for the Socio-Educational Model

Empirical evidence has lent support to the validity of the Socio-Educational Model. Gardner, Lalonde, and Pierson (1983), for example, provided strong support for the internal structure of S-EM. The researchers recruited 140 Canadian freshmen studying L2 French, who filled in a 7-point Likert-scale questionnaire based on the S-EM. The questionnaire assessed 18 variables, such as motivational intensity, French class anxiety, and interest in foreign languages. Upon Linear Structural Relations modelling, the S-EM produced a χ^2 goodness-of-fit value of 254.06 (df=126), and the ratio of χ^2 to df was 1.59. This indicated that overall, the latent and indicator variables measured were adequately consistent with the internal structure of S-EM (Wheaton et al. 1977). Moreover, the researchers found that motivation was predicted by integrativeness (ϕ=0.59) and attitudes towards the learning situation (ϕ=0.32), and motivation, in turn, predicted achievement (ϕ=0.25). This supported a crucial claim in the S-EM, that is, a learner's willingness to integrate into the L2 community and their attitudes towards the learning environment affected their motivation to learn the L2. The validity of the S-EM has also been supported by Gardner (2006; 2010a), Masgoret and Gardner (2003), and Semmar (2006).

In a more recent meta-analysis, it was also found that the constructs in the S-EM (i.e., integrativeness, attitudes toward learning situation, and motivation) were correlated with outcome measures such as grades in the L2 subject, if somewhat weakly. Masgoret and Gardner (2003) conducted a meta-analysis of 75 samples involving 10,489 individuals, and reported that L2 grades were weakly correlated with integrativeness (r=0.24), attitudes towards the learning situation (r=0.24), and motivation (r=0.37). This indicates that each of the constructs in S-EM are somewhat predictive of one's language learning outcomes.

Despite the correlations found, one should also note that some of the correlations were statistically non-significant (p>0.05). The meta-analysis reported that 16 out of 55 findings for integrativeness were non-significant, 2 out of 55 findings for motivation were non-significant, and above all, all 51 findings for attitudes towards the learning situation were non-significant. In the next section, more issues with S-EM constructs will be discussed.

2.2.3. Issues with the Socio-Educational Model

Despite its popularity, S-EM is not without issues. This section discusses theoretical issues related to integrativeness and instrumentality, and methodological issues in the research field.

Firstly, the theorization of "integrativeness" is problematic as it presumes language ownership. According to Gardner, two assumptions of integrative motivation are that (i) "L2 acquisition refers to the development of near-native language skills," and (ii) "such development requires identification with the L2 community" (2001, 2). These assumptions suggest that a language is "owned" by a group of native speakers, and for one to acquire the language successfully, they have to integrate with the native speakers. While this might have been true in Canada, a bilingual context under which the S-EM was developed, this may not apply to other contexts and/or languages. For instance, with the emergence of English as a global language, who are the "native speakers" of English? When one learns English as an L2, are they integrating with the British, Americans, Australians, Singaporeans, or other language groups? Furthermore, S-EM's over-emphasis on integrativeness overlooks cases where learners do not desire to interact with the L2 community, or are unable to interact with the L2 community due to the absence of L2 speakers, but succeed in language learning nonetheless (Taie and Afshari 2015).

Secondly, instrumental orientation remains a weak construct, both theoretically and empirically. In the original framework, Gardner and his colleagues did not extensively

examine the instrumental orientation (Gardner 2010b; McEown, Noels, and Chaffee 2014), as the focus was mostly on integrative orientation. As Gardner (2010a) admitted, and as highlighted by other researchers (Dörnyei 2005; Lamb 2004), the differentiation between integrative and instrumental orientations could be ambiguous in practice. For instance, if one learns English so that they can take up a job in Canada, are they integratively or instrumentally oriented? While Gardner (2010a) argued that the distinction depends on the learner's purpose (i.e., it is "integrative" if they are learning English to engage in the foreign culture, and "instrumental" if they simply wish to complete working tasks), it is totally possible that the learner's learning purpose involves both integrative and instrumental reasons. Additionally, among all the S-EM constructs, instrumentality has consistently been found to have weak correlations with outcome measures. In the meta-analysis conducted by Masgoret and Gardner (2003), it was reported that instrumental orientation had the weakest correlation with grades (r=0.16), objective measures (r=0.04), and self-ratings of achievement (r=0.16) compared with other constructs. This finding was corroborated by Gardner (2005) and Gardner (2006), suggesting that instrumentality may fail to emerge as a construct that can predict one's language achievements.

The last issue pertains to the research field's interpretation and operationalization of the S-EM. As Dörnyei (2005) remarked, many academics have failed to recognize the complexity of S-EM, and falsely reduced the theory to a sum of integrative and instrumental motivation (e.g., Lai 2007). Some researchers have even simplified the S-EM into a false intrinsic-extrinsic dichotomy, which was never mentioned in the original framework (e.g., Wong 2010). In truth, it is difficult to classify integrative/instrumental orientations into intrinsic/extrinsic motivation. Gardner (1985a; 2010a) himself argued that both integrative and instrumental orientations could be extrinsic, as they are oriented towards socio-affective/utilitarian rewards consequential to language learning. Empirical findings on the relationship between integrative/instrumental orientations and intrinsic/extrinsic motivation have been heterogeneous. Pae (2008) argued that integrative orientation was associated with neither intrinsic nor extrinsic motivation, but was nonetheless inclined towards intrinsic motivation. Some empirical research suggests otherwise, stating that integrative orientation strongly associated with intrinsic motivation, or more self-determined forms of extrinsic motivation (McEown, Sawaki, and Harada 2017; Noels 2002). As such, it is difficult to align integrative/instrumental orientations with intrinsic/extrinsic motivation. Whereas simplifying the S-EM into the intrinsic-extrinsic dichotomy might be helpful for certain studies, it

certainly deviates from the original framework. The issue of misrepresentation has been raised in many of Gardner's works (e.g., Gardner 2006; 2007; 2010b), but to this day, the integrative-instrumental dichotomy still pervades many manuscripts submitted to international journals (Dörnyei 2005).

2.3. The Self-Determination Theory (SDT)

2.3.1. Overview of the Self-Determination Theory

Deci and Ryan's (1985) Self-Determination Theory (SDT), which explores various types of intrinsic and extrinsic motives, has been one of the most influential motivational frameworks (Dörnyei 2005). Whereas researchers typically portray intrinsic/extrinsic motivation as a dichotomy (Rigby et al. 1992), the SDT suggests that motivation is a continuum (Figure 2). According to the SDT, there are three types of motivation: intrinsic motivation, extrinsic motivation, and amotivation. Each type of motivation is regulated by autonomous/internal or controlled/external regulation to varying degrees, resulting in six types of motivated behaviours (ranked by degree of "self-determination," from highest to lowest) (R. M. Ryan and Deci 2017) (Figure 3):

1. Intrinsic regulation: Intrinsically motivated and highly autonomous, performed out of interest and enjoyment;
2. Integrated regulation: Extrinsically motivated and highly autonomous, as the value of the extrinsic behaviour is congruent with one's values and beliefs;
3. Identified regulation: Extrinsically motivated and somewhat autonomous, as one has accepted the value of the extrinsic behaviour;
4. Introjected regulation: Extrinsically motivated and somewhat externally regulated, motivated by guilt, shame, contingent self-esteem, or approval from others;
5. External regulation: Extrinsically motivated and externally regulated, directly controlled by external forces such as rewards and punishments;
6. Amotivation: A non-regulated lack of motivation due to a sense of incompetence, lack of interest, or resistance to influence.

To measure the six types of motivated behaviours in L2 learning, Noels et al. (2000) developed the Language Learning Orientations Scale: Intrinsic Motivation, Extrinsic Motivation, and Amotivation. The instrument proved to be reliable and valid, and generated acceptable Cronbach alphas ranging from 0.67 to 0.88.

2.3.2. Support for the Self-Determination Theory

Due to its versatility, the SDT has been widely applied in various contexts, including education, workplace, sports, healthcare, psychotherapy, and religious socialization (R. M. Ryan and Deci 2017). Empirical evidence on SDT and L2 learning have consistently concluded that internally regulated motivated behaviours (i.e., intrinsic regulation and/or identified regulation) are positively correlated with outcome measures like intended effort and engagement (Konno 2011; McEown, Noels, and Chaffee 2014; Noels et al. 2000). In a

study surveying 167 university students, McEown, Noels, and Chaffee (2014) found that intrinsic motivation significantly predicted learners' engagement (β=0.47, p<0.01) and anxiety (β=0.17, p<0.01), and that integrated regulation significantly predicted learners' self-evaluation of their TL competence (β=0.59, p<0.01). However, the more externally regulated forms of motivation (i.e., introjected regulation, external regulation) had significantly negative correlations with criterion measures (e.g., engagement, continuation to learn the L2). This indicates that while intrinsic motivation might encourage engagement and continuation to learn the language, external forces might deter one from doing so.

This claim has been supported by other empirical evidence (e.g., Niemiec and Ryan 2009; R. M. Ryan and Deci 2000; 2013; 2016). Existing empirical research on the SDT generally suggests that the more controlling and externally regulated the learning climate is, the more superficial and less transferable the learning becomes. Furthermore, controlled motivation has been shown to predict greater behavioural problems, disengagement, and even dropout; contrarily, autonomous forms of motivation fostered persistence and higher quality of learning (R. M. Ryan and Deci 2017).

2.3.3. Issues with the Self-Determination Theory

This review identified two issues with the SDT, pertaining to (i) the operationalization of externally regulated motivation and amotivation, and (ii) methodological issues in the field. Firstly, the field has overemphasized the role of intrinsically regulated motivation in L2 learning, leaving the role of externally regulated motivation and amotivation underexplored. As seen from the section above, the field has consistently found negative correlations between externally regulated motivation (i.e., introjected regulation, external regulation, amotivation) and criterion measures (e.g., engagement, continuation to learn the language). For example, McEown, Noels, and Chaffee (2014) found that introjected regulation negatively predicted learner's continuation to learn the language (β=-0.12, p<0.05). This pattern was also observed between external regulation and engagement (β=-0.15, p<0.01), amotivation and engagement (β=-0.17, p<0.01), and amotivation and continuation to learn the language (β=-0.29, p<0.01). This poses two questions concerning the operationalization of constructs: (i) If externally regulated motivation consistently demonstrates a negative relationship with measures of L2 learning, can it still be considered "motivation"? Or would "demotivation" be a more accurate label than "extrinsic motivation"? (ii) Does externally regulated motivation really have no role in promoting L2 learning? Some studies (e.g., Wen

1997) suggested that extrinsic-oriented motivations could also lead to success for language learners; if so, how does externally regulated motivation operate in the L2 learning context?

The second issue pertains to the research field's operationalization of the SDT. While the SDT was designed to be a continuum (R. M. Ryan and Deci 2017), some studies have nonetheless reduced the theory to an intrinsic/extrinsic or an autonomous/controlled dichotomy (e.g., K. W. Lee 2018). While a simplification of the theory could be beneficial to clarity and efficiency in data analysis, it certainly deviates from the original framework, resulting in misrepresentation and false generalizations.

2.4. The Second Language Motivational Self System (L2MSS)

2.4.1. Overview of the L2MSS

Dörnyei (2005) attempted to reframe Gardner's S-EM, with a particular focus on the concept of integrativeness. Dörnyei (2005) identified integrativeness as a psychological and emotional construct, and hence drew on psychological theories to reinterpret the motivational framework. This section outlines the model, followed by a critical evaluation on its constructs and theoretical limitations.

The L2MSS comprises three constructs:

(i) Ideal L2 self (IS), which concerns the qualities one would like to possess as an L2 speaker (e.g., fluently communicating with other English speakers) (Alastair and Liu 2023);

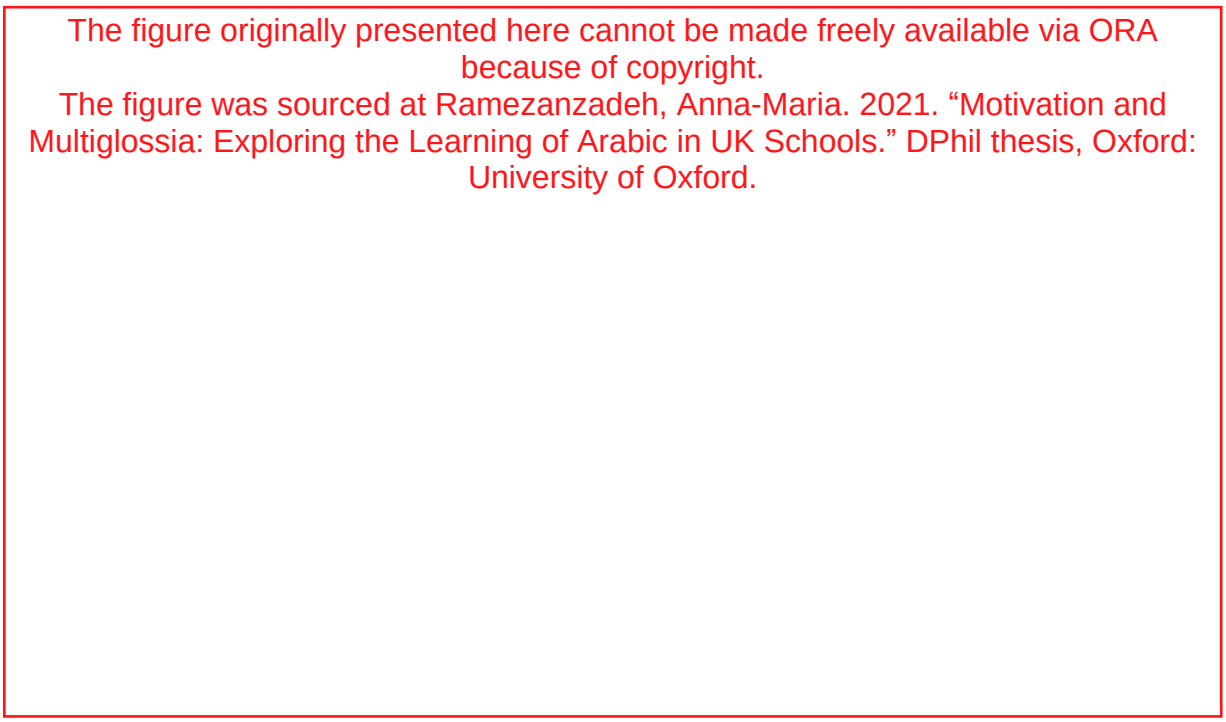
(ii) Ought-to self (OTS), which refers to "the attributes that one believes one ought to possess (i.e., various duties, obligations, or responsibilities) in order to avoid possible negative outcomes" (Dörnyei 2009, 29); and

(iii) L2 learning experience (L2LE), which concerns a learner's "immediate learning environment and experience", such as their teachers, classmates, and textbooks (Dörnyei 2005, 106).

Dörnyei (2005) formulated L2MSS based on two psychological theories: the possible selves theory (Markus and Nurius 1986) and the self-discrepancy theory (E. T. Higgins 1987). The possible selves theory stipulates that one's self-concept comprises images of who they (i) might become, (ii) would like to become, and (iii) are afraid of becoming (Markus and Nurius 1986), which roughly translate to the ideal self (i and ii) and ought-to self (iii) in the L2MSS. Markus and Nurius (1986) argue that these images motivate behavior, as the future images are either to be approached or avoided.

On the other hand, the self-discrepancy theory posits that, when there is a gap between a learner's ideal and current self-concepts, the discrepancy would cause discomfort (E. T. Higgins 1987). This discomfort would then propel one to approach their ideal future selves, i.e., reduce the discrepancy between the actual and desired selves. In E. T. Higgins's (1987) original theory, he distinguished between "own" and "other" sources of IS and OTS, resulting in four future-self constructs: IS-own, IS-other, OTS-own, and OTS-other. OTS-other, for example, would refer to a sense of obligation (OTS) from external sources (other) (e.g., I should study English because my parents expect good grades). This own/other distinction, however, is absent in L2MSS, and remains one of its biggest limitations (more in Section 2.4.3).

Based on these two theories, Dörnyei defined motivation as "the desire to reduce the discrepancy between one's actual and ideal or ought selves" (2005, 100). The conceptualization of the L2MSS is best visualized by Ramezanzadeh's (2021) graph (Figure 4). Since its introduction, the L2MSS has received considerable attention in the research field, surpassing the S-EM to become the most prominent motivational framework in 2013/14 (Boo, Dörnyei, and Ryan 2015). The L2MSS is renowned for its versatility, as it has been widely adopted in both bilingual (e.g., Lamb 2004) and multilingual contexts (e.g., Csizér and Lukács 2010).



2.4.2. Critical Evaluation of the Ideal L2 Self (IS)

Empirical evidence on L2MSS has generally supported the model's validity. The Ideal L2 Self (IS), in particular, has been shown to be a strong predictor of motivation, as measured by Intended Effort. Intended Effort (IE) measures a learner's willingness to take language learning opportunities, as well as the time and effort they intend to spend on L2 learning (Papi et al. 2019). While IE does not have the most straightforward correlation with learning behaviour and achievement (Al-Hoorie 2018; Papi et al. 2019), one should note that objective outcomes such as language achievement are often mediated by factors such as gender, culture, and age (Papi 2010). Moreover, motivation concerns learner's inner desires and language goals, which are mostly internal cognitive/affective processes. In this sense, IE is a sufficient measure of motivation as it measures learners' subjective willingness to engage in L2 learning activities; this may also be one of the reasons why IE has remained a popular outcome measure of L2 motivation (Papi 2010).

In Al-Hoorie's (2018) meta-analysis of 39 L2MSS studies, it was reported that IS was significantly and moderately correlated with IE (r=0.611, p<0.05). This finding was corroborated by various studies conducted around the globe, such as Hong Kong (Dörnyei and Chan 2013), Saudi Arabia (Moskovsky et al. 2016), and Pakistan (Islam, Lamb, and Chambers 2013). Teimouri (2017), for example, investigated a sample of 524 Iranian adolescents studying English as a foreign language. He issued a 6-point Likert scale questionnaire measuring, among other things, the correlation between IS and outcome measures such as IE and willingness to communicate. In the findings, IS was identified as the strongest predictor of IE (β=0.46, p<0.001), and the only predictor of learners' willingness to communicate (β=0.50, p<0.001). This echoes the findings of other research, where IS is found to be a strong predictor of IE, as well as constructs such as motivated learning behaviour (Csizér and Kormos 2009; S. Ryan 2009) and learning styles (Yang and Kim 2011).

Despite IS's strong correlations with subjective measures, its predictive power appears much weaker for objective measures such as L2 achievement. In Al-Hoorie's (2018) meta-analysis, there was no significant relationship between IS and achievement (corrected r=0.103, p>0.05). This was echoed by a series of studies (e.g., Kim and Kim 2014; Lamb 2012). Lamb's (2012) study, for instance, surveyed 527 Indonesian learners aged 13-14, and implemented a C-test to measure their English proficiency. The study reported that IS significantly though weakly correlated with motivated learning behaviour (β=0.25, p<0.05),

but had had no significant correlation with English proficiency (β=0.09, p>0.05). This finding suggests that one's L2 motivation might promote self-regulated learning to some degree, but such learning behaviours might not translate into actual learning outcomes. While this study may seem like counterevidence to the validity of IS, it can also be argued that L2 achievement is mediated by a myriad of factors, ranging from age to motivation to learning styles. This will be discussed later in the meta-analysis.

A more fundamental critique of the IS concerns its theorization. In the original framework, IS differs from OTS by its degree of internalization: while the former focuses on one's internal desires (i.e., I *want to*...), the latter concerns social expectations and responsibilities (i.e., I *should*...) (Dörnyei 2009). However, as Boyatzis and Akrivou (2006) critiqued, it is not always easy to determine whether one's IS is a product of genuine internal desire or that of social expectations. For instance, if a learner states that they want to learn English because "they want to receive better grades at school," is that a desire coming from the self? Or did the learner simply internalize society's expectations for success? Evidently, the theoretical distinction between IS and OTS could be ambiguous, and requires further conceptualization.

2.4.3. Critical Evaluation of the Ought-to Self (OTS)

Compared with the Ideal L2 Self and the L2 Learning Environment, the Ought-to Self (OTS) has been met with less support. For example, Al-Hoorie's (2018) meta-analysis reported that the correlation between OTS and IE (r=0.379) was only half of that of IS (r=0.611) and L2LE (r=0.656). Although some studies reported significant positive correlations between OTS and outcome measures, the links were often weak (e.g., Papi 2010; Taguchi, Magid, and Papi 2009). In Japan, China, and Iran, Taguchi, Magid, and Papi (2009) attempted to validate the L2MSS by surveying 4,943 learners aged 11 to 53, with a mean age of 18.9. They measured, among other constructs, the correlation between learners' OTS and IE, and reported that the relationship was extremely "weak" across contexts (r(Japan)=0.17, r(Iran)=0.12).

In most cases, the relationships between OTS and outcome measures were insignificant (e.g., Lamb 2012), even negative (e.g., Y. Lee and Ahn 2013). For instance, Kormos, Kiddle, and Csizer (2011) surveyed 518 high school students, university students, and young adult learners in Santiago, Chile, measuring their IS, OTS, L2LE, and motivated learning behavior. A multi-group structural equation analysis revealed no significant

relationship between students' OTS and motivated behaviour, regardless of participants' age. This finding was corroborated by Lamb (2012), Papi and Teimouri (2012) and Ramezanzadeh (2021), suggesting that OTS may fail to emerge as a valid construct, both empirically and theoretically.

Furthermore, a more critical limitation of OTS concerns its failure to distinguish between OTS-own and OTS-others, which differentiates the source of the OTS. In Higgin's (1987) self-discrepancy theory, upon which L2MSS was formulated, Higgins classified OTS into "OTS-own" and "OTS-others." Whereas OTS-own is an internal sense of responsibility (e.g., "I should learn English, otherwise I would have problems understanding my grandparents' culture"), OTS-other is rooted in external, social expectations (e.g., "I should learn English to avoid my teacher's reprimand"). In his original theorization of the L2MSS, however, Dörnyei (2005) made no distinction between the two.

As one could easily imagine, OTS-own and OTS-others could at times come into conflict, which creates issues with the theorization of OTS. For instance, while one may be demotivated by their teacher's social pressure (i.e., OTS-other), they could nonetheless feel motivated to learn English as they feel compelled to integrate into their grandparents' culture (i.e., OTS-own). Empirical evidence has provided further evidence on this argument. Blair and Azaz (2019), for instance, investigated the potential difference in OTS-own and OTS-other. The researchers issued questionnaires to 54 American university students studying Arabic as a foreign language, and measured the correlations between OTS-own, OTS-other, and IE. The study reported that whereas OTS-own positively correlated with IE (r=0.10, f^2=0.02), OTS-other negatively correlated with IE (r=-0.26, f^2=0.09). This suggested that OTS-own and OTS-other motivate a person in completely opposite ways: while an internal sense of responsibility could promote one's willingness to engage in L2 learning activities, external obligations could deter one from doing so. This finding was echoed by Teimouri (2017). The weak predictive power of OTS, where it has been detected at all, combined with emerging evidence on the distinction between OTS-own and OTS-other, calls for a need to reconceptualize the OTS.

2.4.4. Critical Evaluation of the L2 Learning Environment (L2LE)

Among the three constructs in L2MSS, the L2 Learning Environment (L2LE) has consistently emerged as the most powerful predictor of outcome measures, including IE (e.g., You, Dörnyei, and Csizér 2016) and learning behaviour (e.g., Lamb 2012). Kong et al. (2018)

surveyed 1,296 Korean university students on the L2MSS constructs, and reported that L2 Attitude (i.e., L2LE) was the strongest factor affecting students' IE, regardless of the L2 they were learning (ϕ(English and Chinese)=0.495, p<0.001; ϕ (Spanish and Arabic)=0.479, p<0.001).

Despite the strong predictive power of L2LE, it is not without limitations. A major critique of L2LE, as Dörnyei (2019) himself admitted, lies in its vagueness and diversity in interpretations. As noted in the example above, L2LE was interpreted as "L2 Attitude" in some studies (e.g., Kong et al. 2018; Kormos, Kiddle, and Csizér 2011), simply "L2LE" in some (e.g., Al-Hoorie 2018; Lamb 2012), and "the extent to which students like learning English" in others (e.g., Csizér and Kormos 2009). While the studies have somewhat unanimously supported the validity of L2LE as a construct, its operationalization is problematic. In his original work, Dörnyei (2005) broadly defined L2LE as "situation-specific motives related to the immediate learning environment and experience" (p.106), but it is unclear what such an "environment" includes (e.g., teachers, peers, course materials), which led to very diverse interpretations of the construct, as exemplified above. In view of this limitation, Dörnyei (2019) recently proposed to reconceptualize L2LE as learners' engagement with (i) the school community, (ii) teaching materials, (iii) learning tasks, (iv) peers, and (v) teachers. However, further empirical research is needed to validate the newly proposed subconstructs, as they were devised merely based on "common sense" (Dörnyei 2019, 25).

2.4.5. Critical Evaluation of the Current Self (CS)

The last issue with the L2MSS pertains to the absence of Current Self (CS). In L2MSS, motivation is defined as "the desire to reduce the discrepancy between one's actual and ideal or ought selves" (p.100). In the model, however, there is no construct to measure one's "actual" (i.e., current) self. It is impossible to study motivation under the L2MSS framework without operationalizing the "discrepancy" in current and future states. Currently, however, the concept of CS is absent from most L2MSS studies (Thorsen, Henry, and Cliffordson 2020), with few exceptions.

One of the studies that incorporated the concept of CS is Hessel (2015). She recruited 97 German university students learning L2 English, and explored the students' perceived self-discrepancy, as well as the internal dynamics of IS. Hessel developed one specific questionnaire item to measure learners' self-discrepancy: "I will be someone who can

converse effortlessly in English," and participants had to rate how closely the statement aligned with their current state. The study reported that self-discrepancy was, in fact, a significant predictor of IE (r=0.49, p<0.01), which highlighted the importance of CS in measuring motivation.

Ramezanzadeh (2021) expounded upon Hessel's (2015) findings by exploring the internal dynamics of CS. She operationalized CS as a tripartite construct, consisting of (i) self-perception, (ii) confidence, and (iii) ability, and calculated self-discrepancy by subtracting CS scores from IS scores. The longitudinal study revealed that self-discrepancy at time point 1 (i.e., at the beginning of the term) did not predict L2 proficiency at time point 2 (i.e., at the end of the term), suggesting that self-discrepancy may not emerge as a significant predictor of L2 achievement. An alternative interpretation is that the relationship between self-discrepancy and outcome measures is not linear, but rather lies on a bell-curve: that is, there exists an optimal point at which self-discrepancy is motivating, before and after which it is not (Ramezanzadeh 2021). As the relationship between self-discrepancy and outcome measures are, potentially, quadratic, it would make sense that self-discrepancy could not significantly predict the outcomes in a linear regression model. Nonetheless, summarizing the findings from these two studies, it is evident that further research is needed to operationalize CS and to investigate whether, and to what extent, self-discrepancy predicts L2 motivation.

2.5. Comparison between the S-EM, SDT, and L2MSS

The sections above outlined the internal dynamics and limitations of the S-EM, SDT, and L2MSS. This section will discuss the comparability of these three models, as all three have been widely adopted in motivational research.

2.5.1. S-EM and SDT

While at first glance, integrative/instrumental orientations from the S-EM might look similar to intrinsic/extrinsic motivations from the SDT, that is not necessarily the case. As illustrated in Section 2.2.3, the alignment of the S-EM and the SDT has been debated in the field. Gardner himself (2010a) argued that both integrative and instrumental orientations could be related to extrinsic motivation, as they are oriented towards socio-affective/utilitarian rewards consequential to language learning. Other scholars have argued otherwise. Noels (2002), for instance, argued that the integrative orientation strongly correlated with more self-determined forms of motivation (i.e., intrinsic motivation and

identified regulation), while instrumental orientation was strongly correlated with external regulation. This was corroborated by McEown, Sawaki, and Harada (2017). Still, other scholars (e.g., Pae 2008) argued that integrative orientation is associated with neither intrinsic nor extrinsic orientation. Given the academic debates, it is evident that large-scale quantitative studies are still required to study the relationship between the S-EM and SDT.

2.5.2. SDT and L2MSS

Regarding the comparability between the SDT and L2MSS, Dörnyei (2009, 30) himself argued that theoretically, intrinsic motivation would align with the L2LE, and extrinsic motivation would align with the OTS. This might be because the "engagement" inherent to L2LE is synonymous with the "interest and enjoyment" in the intrinsic regulation, and that the "extrinsic behaviour" and senses of "guilt, shame, and other-approval" are similar to the sense of responsibility and social expectations highlighted in the OTS.

Dörnyei's claim is supported by Takahashi and Im's (2020) study with 545 freshmen in Japan. Takahashi and Im (2020) collected data on constructs from both SDT and L2MSS, and found that intrinsic motivation had the highest correlation with L2LE (r=0.94, p<0.01), just as Dörnyei proposed. However, as extrinsic motivation is further categorized by its degree of self-determination, Takahashi and Im found that the OTS, OTS-own, and OTS-other were more closely related to introjected regulation (r(OTS)=0.62, p<0.01; r(OTS-own)=0.69, p<0.01; r(OTS-other)=0.54, p<0.01) than external regulation (r(OTS)=0.18, p<0.01; r(OTS-own)=0.13, p<0.01; r(OTS-other)=0.31, p<0.01). The IS, on the other hand, had the highest correlation with identified regulation (r=0.75, p<0.01).

Compared with the aforementioned constructs, intrinsic motivation, integrated regulation, and amotivation still require further explorations on their inter-compatibility. Nonetheless, the existing empirical evidence supports a potential comparability between the SDT and L2MSS. More research is required, however, to support Dörnyei's claim that different L2 motivation theories "converge in a common tripartite construct, which is fully compatible with the L2 Motivational Self-System" (2009, 31).

2.5.3. S-EM and L2MSS

As illustrated above, whereas S-EM focuses on one's external, affective desires to integrate into an L2 community, L2MSS is more internal and cognitive-based (Lamb 2012; Ramezanzadeh 2021). Moreover, in S-EM, learners are perceived to be motivated to learn an

L2 to integrate and identify with another community, while in L2MSS, a learner's motivation stems from future images of themselves (Lamb 2012). As such, S-EM and L2MSS differ from each other by their internal/external orientations and sources of motivation.

Nonetheless, the two models share some similarities, particularly regarding integrativeness and the Ideal L2 Self. According to Dörnyei (2005), L2MSS was an attempt to elaborate on S-EM, particularly the construct of integrativeness. As a result, IS shares some similarities with integrativeness, both on the item level and statistical level. On the item level, the questionnaire items measuring IS and integrativeness are both somewhat communication-orientated. Consider the following examples:

1. "I can imagine myself speaking English with international friends or colleagues" (Papi 2010, item measuring IS).
2. "Studying French can be important to me because it will allow me to be more at ease with fellow Canadians who speak French" (Gardner 1985b, iteam measuring integrative orientation).

In both examples, the statement concerns the learners' future ability to communicate with members of the L2 community. This shows that the instruments measuring IS and integrativeness are more or less similar on the item level. Empirically, various studies have also shown that IS and integrativeness are highly correlated, and that the IS had even higher predictive power than integrativeness for outcome measures (Dörnyei 2009). This finding was corroborated by S. Ryan (2009) and Kim (2009). Kim argued that IS could replace integrativeness as it overcomes the integrative/instrumental dichotomy in existing literature.

Compared with IS and integrativeness, OTS, L2LE, and attitudes toward learning situation are underexplored. On the subscale level, however, it could also be observed that L2LE resembles attitudes toward learning situation. Whereas Dörnyei (2019) proposed to measure L2LE by students' engagement with teachers, teaching materials, the school community, learning tasks, and peers, Gardner's (2001) AMTB also measures attitude through learners' evaluation of teachers and the course. This resemblance highlights a potential comparability between the two constructs, but this similarity is yet to be validated by empirical research.

2.6. Overview of Studies on L2 Motivation and Gender

Empirical studies suggest that gender interacts with motivation; this study investigates the interplay in the Hong Kong secondary school context through a meta-analysis. Existing

literature generally suggests that females exhibit higher motivation in L2 learning, whereas males display lower motivation, irrespective of the target language (Heras and Lasagabaster 2015; Lasagabaster 2016; Powell and Batters 1985). A meta-analysis conducted by Yousefi and Mahmoodi (2022), which investigated 17 independent studies and 18,832 learners, reported that in mixed-sex samples, there is a significant and moderate effect of gender on motivation (k=15, p<0.000, d=0.59). In a more detailed account, Qi's (2021) meta-analysis, consisting of 144 samples and 51,905 participants, analyzed learners' motivation based on the three constructs in the L2MSS: IS, OTS, and L2LE. Qi reported that, whereas there was no evidence that gender significantly impacts the correlation between IE and IS (df=1, Q=0.04, p>0.05) and IE and L2LE (df=1, Q=1.33, p>0.05), it significantly moderated the relationship between OTS and IE (df=1, Q=7.15, p=0.007). In other words, male learners' OTS impacted their IE more than female learners.

Across the world, evidence for the gender difference in motivation is largely consistent. As Dörnyei, Csizér, and Németh remarked, "we do not think that there are many quantitative studies in the L2 literature that examined boys' and girls' attributes or achievement and did not find any salient differences" (2006, 55). They surveyed 13,391 Year 8 pupils in Hungary, collecting data on their integrativeness, instrumentality, attitudes towards the L2 community, vitality of L2 community, cultural interest, milieu, self-confidence, language choice, and intended effort. The researchers reported a consistent tendency for girls' motivation scores to be higher than boys. Out of 120 measures, there were only 22 instances where girls did not outscore boys, and the mean differences sometimes exceeded 0.50 in favour of girls. Their findings are corroborated by other studies conducted in the US (e.g., Bacon and Finnemann 1992; Chavez 2000), Lebanon (e.g., Shaaban and Ghaith 2000), Hong Kong (e.g., Au 2008), and Australia (e.g., Watt 2004), with few cases of counterevidence from the US (e.g., Cortés 2002), Canada (e.g., MacIntyre et al. 2003), and Hong Kong (e.g., Pang 2014).

2.7. Studies on L2 Motivation and Gender in Hong Kong

Despite the global findings, it is essential that gender studies researchers situate their studies within particular cultural contexts. Given that gender is a social construct, gender roles, norms, and expectations vary across cultures and regions (Kalliyath 2015). Gender roles operate very differently in Hong Kong and in Canada, and the gender differences in L2 motivation, likewise, likely also differ by contexts.

The unique sociolinguistic landscape of Hong Kong, a bilingual Asian society, provides the current study with a unique and complex context. With a history of British colonialism, Hong Kong exhibits distinct gender norms and practices that combine both Western biomedical and Chinese cosmological conceptions (i.e., yin/yang) (Ha 2009). It would certainly be intriguing to investigate how such unique gender norms impact students' gender socialization, and hence their motivation in learning "gendered" subjects such as languages. Recent studies reported that English language is still a "gendered" subject in Hong Kong, wherein females receive higher expectations on their English proficiency due to the gender belief that girls are good language learners (Sung 2022).

In Hong Kong, studies on gender differences in L2 motivation are relatively scarce, and the findings have been varied. While some reported a female edge in L2 motivation (Fung 2006; I. Lee, Yu, and Liu 2018; Wong 2010), some reported no significant differences (Au 2008; Pang 2014). Fung (2006), for example, investigated a sample of 20,986 students, and concluded that boys had significantly lower scores in instrumental orientation, integrative orientation, motivational intensity, attitudes towards learning English, and attitudes towards bilingualism. On the other hand, Pang (2014), surveying a sample of 60 freshmen in Hong Kong in a mixed-methods study, reported that gender differences were significant in terms of L2LE, but were insignificant in learners' IS and OTS. Pang argued that gender stereotypes in Hong Kong may lead to gender differences in the construction of IS, which would in turn influence their L2 motivation and L2 achievement. However, the validity of her argument is questionable due to two reasons. First, the study found no significant gender differences in their IS; if, as Pang argued, there are different expectations for male and female learners, and such expectations would affect one's IS, the stereotypes would, in theory, be reflected in IS differences. Second, such "stereotypes" were not reflected in the qualitative analysis of the study. In fact, from the interview transcriptions, it would appear that many participants were against the idea that English was a "female" subject. As such, it seemed that her analysis and conclusion are in contradiction. In Pang's defence, her sample size (n=60) might have been too small to elicit stereotypical views on L2 learning; in larger studies, art subjects such as English have consistently been found to be gendered in Hong Kong (Lai 2007; Zhang 2011).

Given the heterogeneity of findings in Hong Kong, and the importance of studying gender in context, a meta-analysis of gender differences in L2 motivation is much needed in order to provide an overview of the literature and determine whether, overall, there is a relationship. This is a research gap that the current study aims to fill.

2.8. Significance of Adolescence

Empirical studies suggest that gender discrepancies in motivation emerge in adolescence, shedding light onto the importance of this critical period. Some researchers attributed the differences to gender socialization (e.g., Eccles 1987; Watt 2004), arguing that social agents like parents, peers, media, and schools may heighten boys' and girls' interests in doing gender-appropriate activities (e.g., learning languages to be "feminine") particularly during adolescence. The hypothesis of teenage divergence echoed findings from Maccoby and Jacklin's (1974) meta-analysis. The largely-scale analysis, albeit dated, analyzed a sample of 98 studies on verbal development around the world. The researchers found that there were few differences in boys' and girls' verbal abilities from age 3 to 11, but from the ages 10-11 onwards, girls begin to outscore boys at a variety of verbal skills. Maccoby and Jacklin thus argued that there are distinct phases in the development of verbal skills in the two sexes, with "a new phase of differentiation occurring at adolescence" (p.85). This points to a critical phase in language development that is worth scholarly attention, as differences in socialization and motivation during this phase might contribute to boys' and girls' divergence in L2 proficiency, potentially contributing to gender inequality in L2 education.

Adolescence also represents a pivotal juncture in students' life journeys, as it is during this period that career choices are made, and identities are formed (Meeus 2011). Their L2 motivation at this stage plays a central role in shaping their life trajectories, making the issue of gender disparities even more critical. As such, the current study focuses on the secondary school period (ages 12-18), hoping to contribute to the discourse on gender equality in education, especially during this critical period.

2.9. Significance of Study

The heterogeneity in gender and motivation studies, coupled with the unique gender culture in Hong Kong and the criticality of adolescence in language development, underscores the lack of literature summarizing Hong Kong adolescents' gender experiences in L2 learning. These factors provide strong impetus for this meta-analysis. The meta-analysis will synthesize and give an overview of existing studies on the same topic (Popay et al. 2006), which, to my knowledge, has accumulated to a considerable amount with diverse conclusions. The methodology, along with research questions, are outlined in Chapter 3.

3. METHODOLOGY

This systematic review synthesizes empirical evidence on Hong Kong secondary school students' motivation to learn English as Second Language (ESL), and investigates the gender differences in their motivation. The study operationalizes ESL as the learning of English as a language additional to one's mother tongue which, to most Hongkongers, is Cantonese (Census and Statistics Department, HKSAR 2024).

A systematic review is a robust method that utilizes transparent, pre-specified eligibility criteria to synthesize the findings of all research that has addressed similar research questions (J. P. T. Higgins et al. 2023; X. Huang and Chalmers 2023). In the current study, the systematic review was conducted with a meta-analysis, which statistically assesses all available empirical evidence on the given topic. The meta-analysis was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al. 2021), which includes 27 items to guide the reporting of systematic reviews. The *Cochrane Handbook for Systematic Reviews of Interventions* (J. P. T. Higgins et al. 2023), a guidebook written by an international network of systematic review professionals, was also consulted for risk of bias assessments and statistical analysis.

In this chapter, the review process and justifications for my methodological choices will be clearly stated such that the review is transparent and replicable.

3.1. Research Questions

1. What is the extent and nature of empirical research into gender and motivation towards learning L2 English in Hong Kong secondary schools?
2. What is the relationship between gender and Hong Kong secondary students' L2 learning motivation?
3. To what extent do other variables (e.g., age, school banding) mediate any relationship between gender and L2 motivation?
4. How reliable is the existing empirical evidence on gender and L2 motivation in Hong Kong?

3.2. Protocol Registration

The protocol for this systematic review was prospectively registered on the International Database of Education Systematic Reviews (IDESR), and can be accessed via https://idesr.org/article/IDESR000118 (accessed on 23 July 2024).

3.3. Search Strategy

3.3.1. List of Databases

Chinese and English databases covering fields of education, psychology, and social science were searched for this systematic review (see Table 2). The following list of academic databases was compiled based on the recommendations from the experienced librarian at the Department of Education, University of Oxford. Furthermore, Google Scholar was also used to identify additional citations. The Chinese and English search strings were applied to the Google Scholars search engine, and the first five pages of results were obtained for screening. The last date of my initial search was 29 April 2024.

While in the protocol, I intended to search ProceedingsFirst for conference papers, it was later discovered that ProceedingsFirst only indicated the titles and authors of conference papers, but did not include the abstracts. This made the title and abstract screening impossible for this database. ProceedingsFirst was hence excluded from the initial search. As such, the review only relied on the Oxford University Research Archive and PapersFirst to search for conference proceedings.

Table 2. List of Databases

Category	Databases	Date last searched
General	• Airiti Library	2024/02/14
	• China Knowledge Resource Integrated China Journal Full-text Database (CNKI CJFD)	2024/02/14
	• Hong Kong Journals Online	2024/02/14
	• SCOPUS	2024/02/14
Conference Proceedings	• Oxford University Research Archive	2024/02/14
	• PapersFirst	2024/02/14
Education	• Australia Education Index	2024/02/14
	• British Education Index	2024/02/27
	• ProQuest (including ERIC, Linguistics collection, LLBA, ProQuest Social Science Premium Collection)	2024/02/20
Theses	• CityU Scholars (Student Theses)	2024/04/24

	• CUHK LibrarySearch (CUHK Theses)	2024/02/29
	• EdUHK Research Repository	2024/02/29
	• HKBU Scholars (Student theses)	2024/02/29
	• HKUST Electronic Theses	2024/04/24
	• Lingnan University Theses and Dissertations	2024/04/24
	• PolyU Electronic Theses	2024/04/24
	• ProQuest Dissertations & Theses Global	2024/04/24
	• The HKU Scholars Hub	2024/04/29
	• Theses and Dissertations (HKMU)	2024/02/29
Psychology	• PsycINFO	2024/02/29
Science	• Web of Science Core Collection	2024/02/29

3.3.2. Search String

Upon discussion with my supervisor and the librarian, it was agreed that my search string should cover four concepts: i) motivation, ii) Hong Kong, iii) secondary school students, and iv) English as a second language. Search terms were formulated based on each concept and refined by extensive piloting in databases. The search terms are then connected by Boolean operators (e.g., AND) to form search strings.

Two search strings were formulated to cover literature in both English and Chinese. An English search string was first constructed; it was then translated into Chinese and reviewed by Chinese-speaking postgraduate applied linguistics students at Oxford to ensure accuracy and coverage of all possible translations. Extensive piloting of both search strings was also conducted.

One of the major challenges in search string construction lies in the extensive terminology in motivation literature. As Murphy and Alexander (2000) noted, motivation researchers use synonyms to denote varied constructs, or different terminologies to reference the same construct. For instance, while some studies may use “future self” to denote learners’ goals, it is termed “ideal self” in other research. To ensure all relevant literature is included in my search, I piloted my search terms on SOLO and identified the keywords of relevant motivation research. I then modified my search terms to include all relevant keywords in these studies.

As such, the English and Chinese search strings are as follows:

(Motivat* OR efficacy OR ambition OR attitud* OR L2MSS OR “second language motivational self system” OR “L2 motivational self system” OR “self-concept” OR “self concept” OR “L2 self” OR “ideal self” OR “ought to self” OR “ought-to self” OR “possible self” OR “current self” OR “future self” OR incentive*)
AND (gender OR sex OR female OR male OR woman OR women OR man OR men OR boy* OR girl*)
AND (“Hong Kong” OR “HK”)
AND (secondary OR adolescen* OR teenag* OR HKDSE OR DSE OR “Diploma of Secondary Education”)
AND (“English as a second language” OR ESL OR “second language” OR L2 OR “English as a foreign language” OR EFL OR “English as an additional language” OR EAL OR “second language acquisition” OR SLA OR English OR “English language” OR “foreign language” OR FL OR “additional language”)

(“動力” OR “動機” OR “動因” OR “自我效能” OR “興趣” OR “志向” OR “態度” OR “二語動機自我系統” OR “第二語言自我形象動機” OR “意識” OR “自我” OR “二語自我” OR “理想我” OR “必須我” OR “可能自我” OR “目前我” OR “未來我” OR “誘因”)
AND (“性別” OR “性/別” OR “女” OR “男”)
AND (“香港”)
AND (“中學” OR “高中” OR “初中” OR “青年” OR “青少年” OR “年輕” OR “香港中學文憑考試” OR “文憑試” OR “DSE” OR “HKDSE”)
AND (“英語” OR “英文” OR “二語” OR “外語” OR “二外” OR “第二語言” OR “語言學習” OR “二語學習” OR “語言習得” OR “二語習得” OR “英語教學” OR “語文教學” OR “語言輸入” OR “L2” OR “ESL” OR “EFL” OR “EAL”)

To increase the accuracy of the search, the search terms were applied everywhere except in the main text (i.e., in title, abstract, and keywords). In rare instances where advanced search is unavailable (i.e., in Hong Kong Journals Online, HKBU Scholars, HKUST Electronic Theses, HKU Scholars Hub), a simple search was conducted such that the search string was applied everywhere, including the full text.

3.4. Eligibility Criteria

The studies obtained were screened based on the following criteria (see Table 3):

Table 3. Eligibility Criteria

Criteria	Inclusion	Exclusion	Rationale
Accessibility	Studies that are physically or digitally accessible, including all literature that can be physically borrowed from Bodleian/Hong Kong libraries.	Studies that are physically and digitally inaccessible.	To obtain as much literature as possible. However, some might be inaccessible due to geographical limitations.
Publication type	All primary empirical studies regardless of publication status (including dissertations, conference papers, and grey literature).	Secondary studies (e.g., literature review).	The inclusion of grey literature mitigates publication bias. Drawing from empirical studies ensures the possibility of a statistical analysis of findings.
Language	Studies written in English, traditional Chinese, or simplified Chinese.	Studies that are written in neither English nor Chinese.	Studies written in Chinese (both traditional and simplified) are included to expand potential findings. Given that English and traditional Chinese are the two major written languages used in Hong Kong, it is likely that the

			majority of research published relating to this area will be captured. Studies published in other languages fall outside of the primary investigator's linguistic repertoire.
Context of Study	Studies involving learners learning L2 English in Hong Kong.	Studies involving learners learning L2 English outside Hong Kong (e.g., studies involving Hong Kong students in Canada).	Studies should focus on learners in Hong Kong to ensure comparability of social and cultural contexts.
Study design	Quantitative or mixed-methods studies.	Studies that only collected qualitative data.	To ensure the possibility of a statistical analysis of findings.
Focus of study	Research that studied learners' motivation towards learning English as a Second Language.	Research that only studied learners' general academic motivation, but did not specify their motivation towards L2 learning.	Students' motivation for learning may vary across subjects. By limiting the research to L2 motivation, it ensures comparability and relevance to the research question.
Gender composition	Studies that involved both male and female participants.	Studies that only included one gender.	To ensure that an effect size of gender can be calculated.
Participants' education levels	Studies with participants studying in any grade from Secondary 1 to Secondary 6 (ages 12-18).	Studies wherein none of the participants studied in any grade ranging from Secondary 1 to	The secondary school period corresponds to adolescence (ages 12-18), which the current study concerns.

		Secondary 6, or wherein the age of participants is not specified.	
Participants' ethnicity	Studies that examined local students in Hong Kong, including ethnic minorities such as Pakistanis who were born and raised in Hong Kong.	Studies that exclusively examined non-local students pursuing their studies in Hong Kong (e.g., Turkish international students' learning experiences in Hong Kong).	To ensure a level of comparability between participants' social backgrounds.
Sample size	Studies that reported the total sample size, male sample size, and female sample size.	Studies where the total sample size, male sample size, and female sample size cannot be determined.	To ensure that an effect size of gender and the standard error can be calculated.
Outcomes with scores	Studies that reported the respective motivation mean scores and SD for male and female participants.	Studies that did not report the respective means and SD in male and female groups.	To ensure that an effect size of gender and the standard error can be calculated.

3.5. Data Management

Returned records from the databases were uploaded to Rayyan, a free AI-powered tool designed for systematic reviews. The software was used for duplicate removal, title and abstract screening, and full-text screening. Upon retrieval of eligible full texts, data extraction and risk of bias assessments were performed using Microsoft Excel spreadsheets. Detailed notes documenting the database search and screening process were kept in a Microsoft Excel research journal.

3.6. Selection Process

3.6.1. Title and Abstract Screening

Upon the removal of duplicate records via Rayyan, all records were screened by their title and abstract against the inclusion and exclusion criteria. At this stage, any records that clearly violated the eligibility criteria were excluded, while the rest were retained for full-text screening. Two independent reviewers, recruited through personal connections and trained in applied linguistics, reviewed a random sample of 5% (n=111) of the titles and abstracts. The 5% percentage was determined by a recent educational systematic review conducted by Huang and Chalmers (2023), and was deemed appropriate for this study due to the time constraint and high volume of records.

Before the screening, a target inter-rater agreement coefficient was set at 90% (Ramezanzadeh and Woore 2023). Since the inter-rater agreement coefficient reached 97%, I proceeded to screen the remaining titles and abstracts.

3.6.2. Full-Text Screening

In the full-text screening, I retrieved and read through all the studies and decided their suitability for the review based on the inclusion and exclusion criteria. Due to the low volume of full-texts to be screened, the two independent reviewers reviewed a random sample of 30% (n=17) of retrieved full-texts (X. Huang and Chalmers 2023).

The inter-rater agreement for the first screen was only 82%, which was lower than the target. Following the same procedure as the title and abstract screening, all disputes over studies' eligibility were resolved through discussion, and the reviewers repeated the process for a further 30% of the full-texts. In the second screen, the inter-rater agreement reached 100%, and I proceeded to screen the remaining full-texts. All reasons for exclusion were recorded on Rayyan.

3.7. Data Extraction

A data extraction sheet, modified from Z. Huang (2020), X. Huang and Chalmers (2023), and J. P. T. Higgins et al. (2023), was used to record information from eligible studies. The extracted data were recorded in an Excel spreadsheet to streamline data management. The data collected include participant demographics, study design, and assessment tools used. For the purpose of this study, fields were added to examine the motivational framework applied (where stated) and the components of motivation measured (e.g., orientations; self-concepts). The following data items were extracted (see Appendix A for detailed list):

- Bibliographic information: Reference citation, year of publication, gender of first author, database/information source, publication type, language of publication, source of funding.
- Setting of study: Context of study.
- Theoretical framework: motivational theory applied (if applicable), motivational components examined.
- Methods: Research questions, study design (cross-sectional/longitudinal; study duration, if applicable), intervention (if applicable), comparator (if applicable), measurement tools.
- Participants: Recruitment method, gender composition, demographics/eligibility criteria, sample size, attrition.
- Outcomes: ESL skill(s) assessed, aspect(s) of motivation assessed, outcomes with scores, effect size.
- Limitations: author acknowledged, researcher observed.

3.8. Quality Assessment

3.8.1. Risk of Bias in Individual Studies

This study adopts the Quality Assessment with Diverse Studies (QuADS) (Harrison et al. 2021), which is a revision of the Quality Assessment Tool for Studies with Diverse Designs (QATSDD). QATSDD has been widely adopted in systematic reviews but has been criticized for its vague assessment criteria and quantitative bias (Harrison et al. 2021). QuADS overcomes said weaknesses by adding explicit examples to each descriptor and revising the language to accommodate both quantitative and qualitative methodologies.

QuADS is a quality assessment tool suitable for mixed or multi-methods studies (Harrison et al. 2021), and was deemed appropriate for this systematic review due to the emergence of mixed-methods motivation studies over the past decade (Boo, Dörnyei, and Ryan 2015). QuADS evaluates the trustworthiness of each study based on 13 descriptors, each rated on a 4-point scale (0-3) (see Appendix B for full rubrics). The 13 descriptors include: theoretical or conceptual underpinning to the research, research aim(s), research setting and target population, study design, sampling method(s), data collection tool(s), data collection procedure, recruitment method, analytic method, consideration research stakeholders, and evaluation of strengths and limitations. The face and content validity of QuADS has been verified by 10 researchers specializing in systematic reviews, generating a substantial inter-rater reliability (k=0.66) (Harrison et al. 2021).

As with full-text screening, two independent reviewers conducted 30% (n=3) inter-rating of risk of bias assessments. The inter-rater agreement coefficient was 92.3%.

3.8.2. Risk of Bias across Studies

To rate the degree of confidence in cumulative evidence, the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system was adopted. Although initially designed for clinical studies, GRADE is useful for evaluating the quality of cumulative evidence in systematic reviews as it provides detailed and explicit criteria for ratings of quality (Guyatt et al. 2008).

The GRADE system classifies the quality of evidence into four levels: high, moderate, low, or very low. The higher the quality of evidence, the more confident we can be that the true effect lies close to that of the estimated effect (Balshem et al. 2011). In systematic reviews, the body of eligible literature begin with the assumption of "high" quality (J. P. T. Higgins et al. 2023). The quality of evidence is then upgraded or downgraded based on five downgrading factors and three upgrading factors. Downgrading factors include: risk of bias in individual studies, inconsistency in findings across studies, indirectness of evidence (i.e., heterogeneity of the body of literature), imprecision (i.e., wide confidence intervals), and publication bias. Upgrading factors include: large effect sizes, high dose-response gradient (i.e., effectiveness of interventions), and likeliness of unmeasured confounding variables to underestimate the actual effect size. See Appendix C for a GRADE evaluation rubric, which was adapted from the GRADE guidelines (Guyatt, Oxman, Kunz, Brozek, et al. 2011; Guyatt, Oxman, Kunz, Woodcock, Brozek, Helfand, Alonso-Coello, Glasziou, et al.

2011; Guyatt, Oxman, Kunz, Woodcock, Brozek, Helfand, Alonso-Coello, Falck-Ytter, et al. 2011; Guyatt, Oxman, Montori, et al. 2011; Guyatt et al. 2013).

Given the inherent subjectivity of the GRADE evaluation, two independent reviewers evaluated the degree of confidence in cumulative evidence using GRADE. The inter-rater agreement coefficient was 100%, and the overall confidence was rated "very weak".

3.9. Data Synthesis

Given sufficient homogeneity of studies' measures of motivation, a statistical analysis was carried out. As in line with standard meta-analysis procedures (J. P. T. Higgins et al. 2023), the statistical analysis assumes normal distribution in each outcome of each study, which allows for parametric tests.

For each relevant measure in the eligible studies, an effect size (Cohen's *d*) was computed based on the available information. The effect direction was standardized such that a positive value indicates a male advantage, and a negative one indicates a female advantage. The *Q*-value and I^2 index were then computed to measure the heterogeneity of all effect sizes, testing the null hypothesis that the variability in data is caused by chance rather than heterogeneity (J. P. T. Higgins et al. 2023). Given the heterogeneity in effect sizes ($Q(7)=236.178$, $p<0.001$, $I^2=0.972$), a random-effects model was applied to accommodate between-studies variances (Borenstein, Hedges, and Rothstein 2007).

To address RQ1, the publication years, study designs, sample sizes, outcome measures, and motivation theory applied in eligible literature were summarized. To address RQ2, this review reported the sample size, male and female means of L2 motivation, standard deviations, *p*-value, Cohen's *d*, and standard error across the body of eligible literature. A forest plot was then employed to visualize the effect sizes across studies and the combined effect size, presented with a 95% confidence interval. To address RQ3, a narrative synthesis was carried out as there was insufficient data to conduct a metaregression for different moderators. A narrative synthesis, combined with statistical data, was hence used to organize, describe, and interpret the moderators of study findings (Thomson and Campbell 2020). To address RQ4, the risk of bias within and across studies was reported within the QuADS and GRADE frameworks. Publication bias was also reported with Egger's regression.

4. FINDINGS

Figure 5 shows the screening process. The initial search resulted in 2,487 records, including 2,387 from databases and 100 from Google Scholars. Upon removing 274 duplicates, 2,213 records were eligible for title and abstract screening, after which 2,158 were removed. 55 full-texts were retrieved for full-text screening. 47 were excluded for the wrong focus of study (i.e., irrelevance) (n=25), wrong outcomes (n=13), wrong population (n=6), wrong publication type (n=2), and wrong study design (n=1). The search resulted in eight eligible studies:

1. Au, Mei Yan Florence. 2008. "Investigating Gender in Students' English Learning Beliefs in an English as a Second Language (ESL) Class." MA thesis, Hong Kong: The University of Hong Kong. https://doi.org/10.5353/th_b4126205.
2. Chan, Tat Wah Edmond. 2000. "A Case Study on the Motivational Pattern of Learning ESL in S.W.C.S. Chan Pak Sha School: In Application to Self Access Learning." MA

thesis, Hong Kong: Hong Kong Baptist University. https://scholars.hkbu.edu.hk/en/studentTheses/a-case-study-on-the-motivational-pattern-of-learning-esl-in-swcs-.

3. Fung, Kam Yin. 2006. "Sex Differences in English Learning in Junior Secondary School in Hong Kong." EdD Thesis, Hong Kong: The Chinese University of Hong Kong. https://repository.lib.cuhk.edu.hk/en/item/cuhk-343847.
4. Lai, Mee-Ling. 2007. "Gender and Language Attitudes: A Case of Postcolonial Hong Kong." *International Journal of Multilingualism* 4 (2): 83–116. https://doi.org/10.2167/ijm068.0.
5. Lee, Icy, Shulin Yu, and Yuan Liu. 2018. "Hong Kong Secondary Students' Motivation in EFL Writing: A Survey Study." *TESOL Quarterly* 52 (1): 176–87. https://doi.org/10.1002/tesq.364.
6. Lee, Kwok Wai. 2018. "Reading Motivation in Chinese and English in Hong Kong Secondary Schools Students: Influence of Schools and Demographic Factors." EdD Thesis, Hong Kong: The Education University of Hong Kong. https://repository.eduhk.hk/en/publications/reading-motivation-in-chinese-and-english-in-hong-kong-secondary-.
7. Wong, Ruth. 2010. "Carrot or Stick? An Investigation into Motivation Orientations in Learning English among Hong Kong Chinese Students." *Revista Brasileira de Linguística Aplicada* 10 (1): 71–87. https://doi.org/10.1590/S1984-63982010000100005.
8. Yung, Kevin Wai-Ho. 2019. "Exploring the L2 Selves of Senior Secondary Students in English Private Tutoring in Hong Kong." *System* 80 (February):120–33. https://doi.org/10.1016/j.system.2018.11.003.

4.2. Description of study characteristics

This meta-analysis starts with a summary of the characteristics of eligible studies, as advised by J. P. T. Higgins et al. (2023). Table 4 summarizes the references, publication years, sample sizes, age groups, medium of instruction, school banding, study design, and the motivational theory applied in each study. The full data extraction sheets, including the background and findings of each eligible study, are available in Appendix D.

4.2.1. Publication Type

Figure 6 shows that half of the eligible studies were journal articles (*n*=4), while the rest were an even split between doctoral theses (*n*=2) and masters' theses (*n*=2).

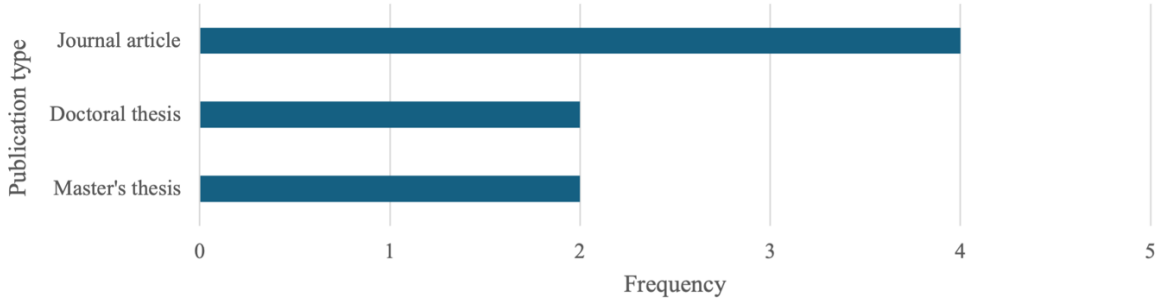


Figure 6. Frequency of publication types

4.2.2. Publication Year and Study Design

The studies predominantly applied mixed methods (*n*=5), while three employed purely quantitative methods (Figure 7). As shown in Figure 8, half of the eligible studies were conducted in the 2000s, with four conducted more recently, in 2010, 2018 and 2019. There is no obvious methodological trend observed (Figure 8).

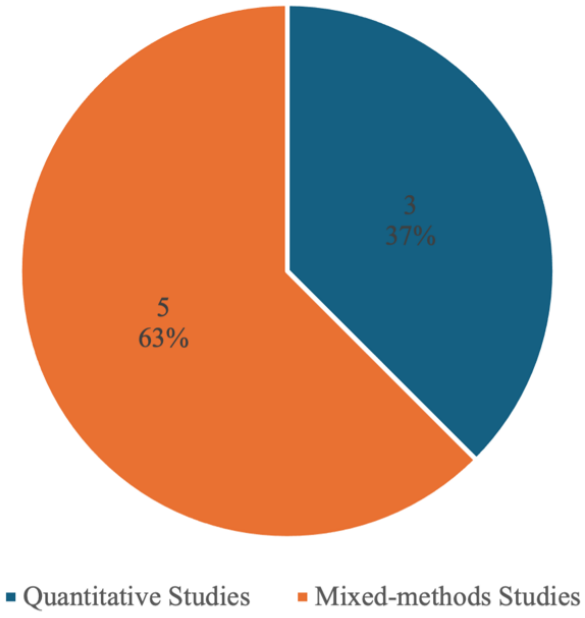


Figure 7. Frequency and percentage of study design

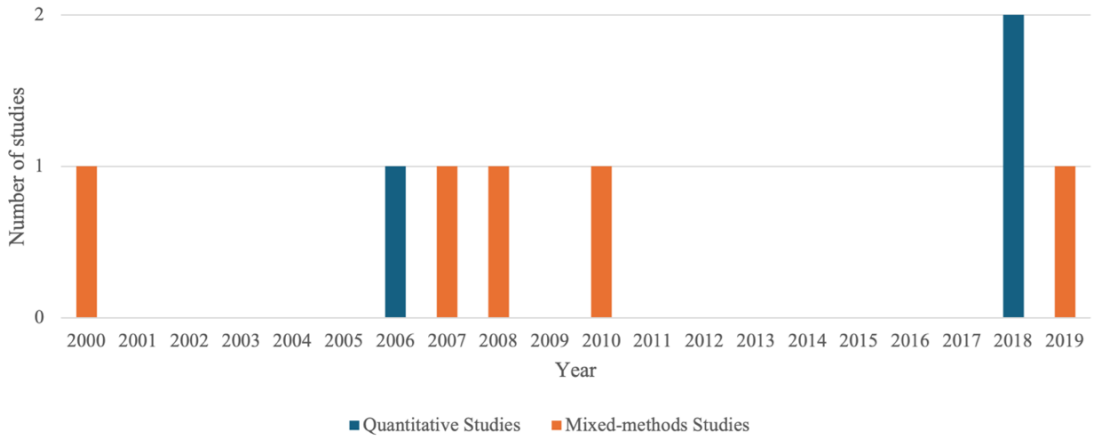


Figure 8. Number of studies by publication year and study design

4.2.3. Age Groups and Sample Size

Of all the age groups studied, Secondary 3 students (*n*=21,511) dwarfed the rest (Figure 9), as Fung's (2006) large-scale study analyzed a sample of 20,986 Secondary 3 students. On the other hand, Secondary 2 students (*n*=41) are the least studied group, as they were only involved in one study (i.e., Au 2008). In summary, Secondary 3 students made up 87.230% of the total sample, followed by Secondary 4 (5.827%), Secondary 1 (2.753%), Secondary 5 (2.165%), Secondary 6 (1.857%), and Secondary 2 (0.166%).

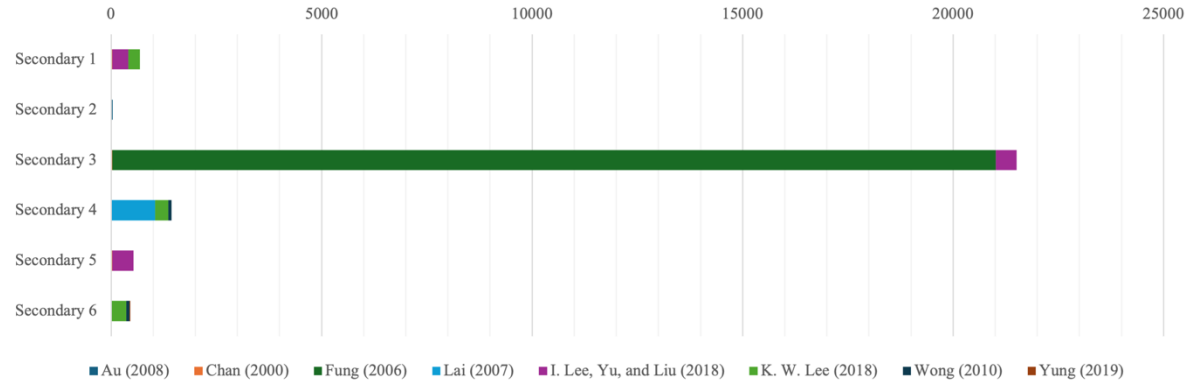


Figure 9. Age groups by sample size

4.2.4. School Banding

In Hong Kong, secondary schools are divided into three bands: Band 1, Band 2, and Band 3. Band 1 represents high-achieving schools, Band 2 represents medium-achieving schools, and Band 3 represents low-achieving schools (Chung 2016). The banding of schools is determined by their students' academic achievements: it considers students' examination results in Primary 5 and 6, as well as their performance in the territory-wide Pre-Secondary One Hong Kong Attainment Test (Ho and Lee 2020). Since research has revealed correlations

between one's academic achievements and L2 motivation (e.g., Dörnyei and Chan 2013), this paper hypothesizes that school banding could moderate potential differences in L2 motivation. The school bandings of participants are hence extracted and listed in Table 4.

In Chan (2000), it was mentioned that the students belonged to "Band 2-3" (p.25). The participants from this study were hence coded as Band 2.5 students. As shown in Figure 10 and Figure 11, the school banding of most participants is unknown as the studies did not report relevant data (90.057%). For participants with known school bandings, most were from Band 1 schools (44.821%), followed by Band 2 (27.121%), Band 3 (25%), and Band 2.5 (3.059%).

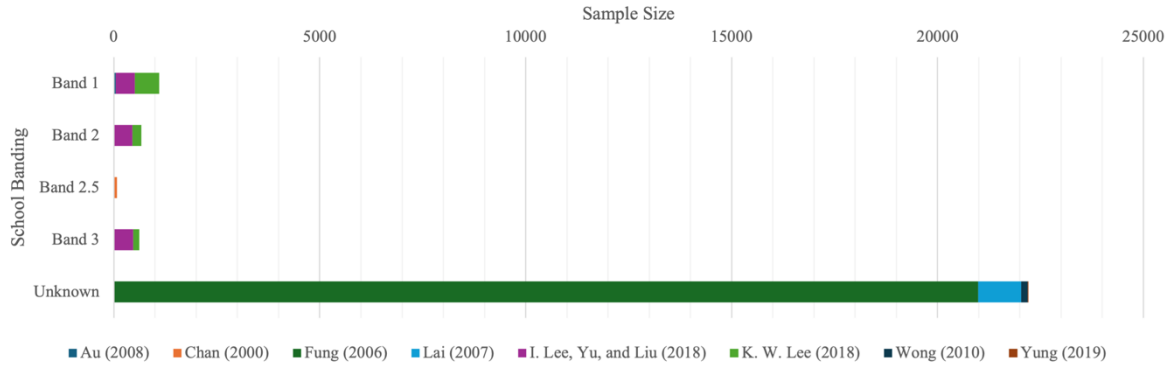


Figure 10. School banding by sample size

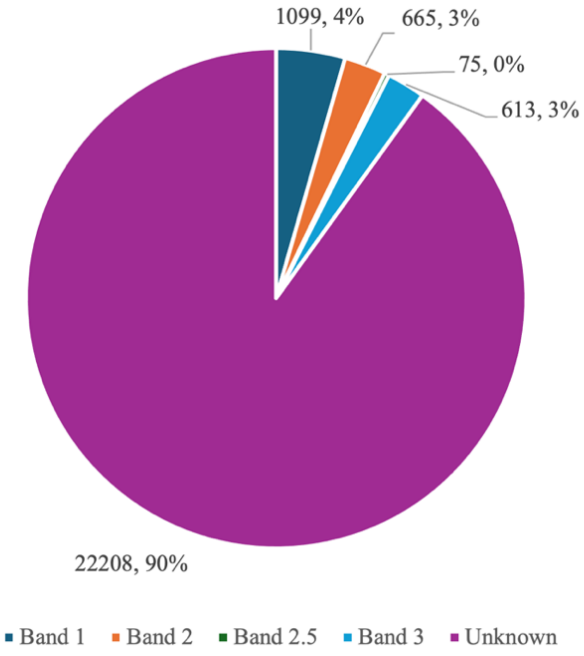


Figure 11. Frequency and percentage of participants from different school bandings

4.2.5. Motivational Theories Applied

Regarding the motivational theories applied, there is an even split between S-EM (*n*=3), L2MSS (*n*=2), and SDT (*n*=2) (Figure 12). Au (2008) did not specify the motivational theory it applied, but judging from its questionnaire items, it appears to align with SEM as it concerns participants' integrative orientation (e.g., "I would like to speak English well so that I can know more about English-speaking people or even make friends with them.") and instrumental orientation ("If I learn to use English very well, I will have better chances to receive better education or even get a good job.").

Despite the motivational theories identified, the outcome measures in some of the eligible studies were questionable, as they did not align with the theories they claimed to have employed (Figure 13). This will be explored in the discussion.

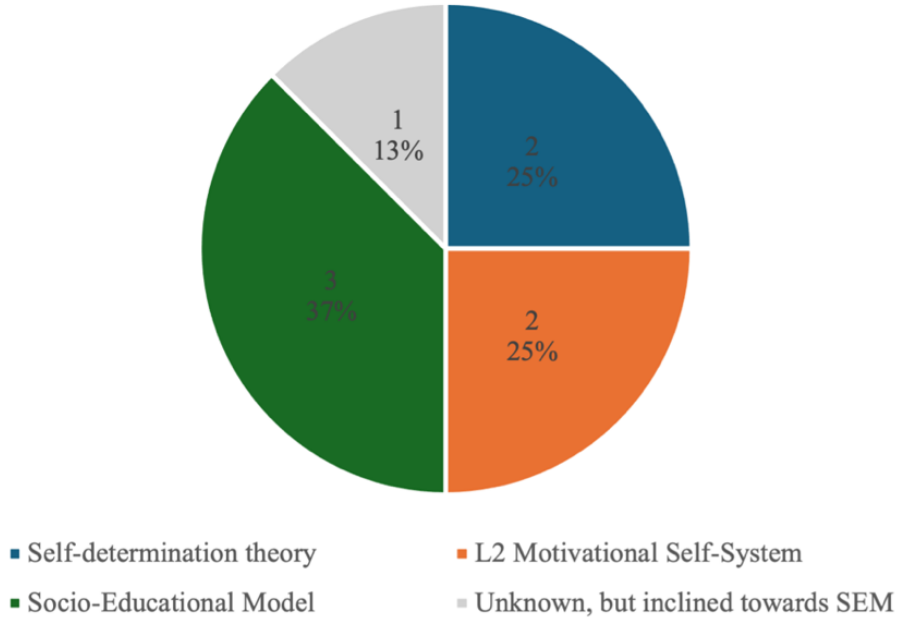


Figure 12. Frequency and percentage of motivational theories applied

OUTCOME MEASURES	SOCIO-EDUCATIONAL MODEL	SELF-DETERMINATION THEORY	L2MSS
MODIFIED AMTB*	✓		
INTEGRATIVE/INSTRUMENTAL ORIENTATION	✓		
INTRINSIC/EXTRINSIC MOTIVATION	✓	✓	✓
CONTROLLED/AUTONOMOUS READING MOTIVATION		✓	
IDEAL SELF, OUGHT-TO SELF, AND OTHER MEASURES OF WRITING MOTIVATION**			✓

*Measures in modified AMTB: (i) integrative/instrumental orientation, (ii) attitudes towards English and bilingualism, (iii) motivational intensity, and (iv) desire to learn.
** Measures of writing motivation include: (i) importance of writing, (ii) interest in writing, (iii) self-regulation, (iv) writing efficacy, and (v) classroom writing context.

Figure 13. Outcome measures by theories applied

Table 4. Summary of the background information of eligible studies

Reference	Type	Sample size	Age groups	Medium of instruction	School banding	Study design	Motivational theory applied
Au (2008)	Master's thesis	41 (20 males and 21 females)	Secondary 2	English	Band 1	Mixed-methods; cross-sectional	Unknown, but similar to the Socio-Educational Model based on my judgement.
Chan (2000)	Master's thesis	75 (41 males and 34 females)	Secondary 1 (*n*=25), Secondary 2 (*n*=25), Secondary 3 (*n*=25)	Chinese for Secondary 1 students; English for Secondary 2-3 students.	Band 2-3	Mixed-methods; cross-sectional	Self-Determination theory
Fung (2006)	Doctoral thesis	20,986 (11,052 males and 9,934 females)	Secondary 3	N/A	N/A	Quantitative; cross-sectional	Socio-Educational Model

Reference	Type	Sample size	Age groups	Medium of instruction	School banding	Study design	Motivational theory applied
Lai (2007)	Journal article	1,048 (555 males and 493 females)	Secondary 4	Chinese (*n*=564); English (*n*=448)	N/A	Mixed-methods; cross-sectional	Socio-Educational Model
I. Lee, Yu, and Liu (2018)	Journal article	1,395 (696 males and 699 females)	Secondary 1 (*n*=386), Secondary 3 (*n*=500), Secondary 5 (*n*=509)	N/A	Band 1 (*n*=471), Band 2 (*n*=452), Band 3 (*n*=472)	Quantitative; cross-sectional	L2MSS
K. W. Lee (2018)	Doctoral thesis	941 (426 males and 515 females)	Secondary 1 (*n*=268), Secondary 4 (*n*=310), Secondary 6 (*n*=363)	N/A	Band 1 (*n*=587), Band 2 (*n*=213), Band 3 (*n*=141)	Quantitative; cross-sectional	Self-Determination theory
Wong (2010)	Journal article	156 (74 males and 82 females)	Secondary 4 (*n*=79), Secondary 6 (*n*=77)	N/A	N/A	Mixed-methods; cross-sectional	Socio-Educational model

Reference	Type	Sample size	Age groups	Medium of instruction	School banding	Study design	Motivational theory applied
Yung (2019)	Journal article	18 (9 males and 9 females)	Secondary 6	Chinese (*n*=9); English (*n*=9)	N/A	Mixed-methods; cross-sectional	L2MSS

4.3. Study Findings

Table 5 summarizes the outcome measures, measurement tools, male and female means of L2 motivation, standard deviations (SD), Cohen's *d*, *p*-value, and standard error of the mean (SEM) across the eight eligible studies. The figures were computed based on the data provided in the original papers, unless otherwise specified.

In cases where the standard deviations were reported separately (e.g., the study reported the male and female means of L2 motivation, but not the overall mean), the SDs were combined using the following formula, as advised by J. P. T. Higgins et al. (2023):

$$s_z^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-1} + \frac{nm(\bar{x}-\bar{y})^2}{(n+m)(n+m-1)}$$

Among the eight eligible studies, Fung (2006) (n=20984, p<0.000, d=-0.100), Lai (2007) (n=1048, p<0.000, d=-0.293), and I. Lee, Yu, and Liu (2018) (n=1395, p=0.004, d=-0.049) found significant gender differences between students' L2 motivation. Figure 14 visualizes the effect sizes of all eligible studies at a 95% confidence interval, and found that the combined effect size across studies was -0.161 (p=0.00, d=-0.161). This suggests that the gender differences observed across studies are significant, but so small that it can be considered negligible (Cohen 1988).

It should be noted, however, that two of the eligible studies studied participants' motivation in specific ESL skills: K. W. Lee (2018) studied participants' English reading motivation, while I. Lee, Yu, and Liu (2018) studied participants' English writing motivation. When analysed separately, the studies on specific ESL skills found a significant but negligible gender difference (p=0.04, d=-0.037) (Figure 16), whereas the studies concerning general L2 motivation found a significant, small effect size in favour of females (p=0.00, d=-0.226) (Figure 15). This suggests that in general, female secondary school students in Hong Kong are slightly more motivated than their male counterparts in learning ESL, but the effect may not be observed when focused on specific language skills.

Furthermore, taking publication years into account, a regression analysis at the 95% confidence interval found no significant correlation between the publication year and observed effect size (t=-1.071, F=1.126, B=-11.145, p=0.325). This suggests that in the eligible studies, the observed gender gap in L2 motivation (if any) has neither significantly narrowed nor widened from 2000 to 2019.

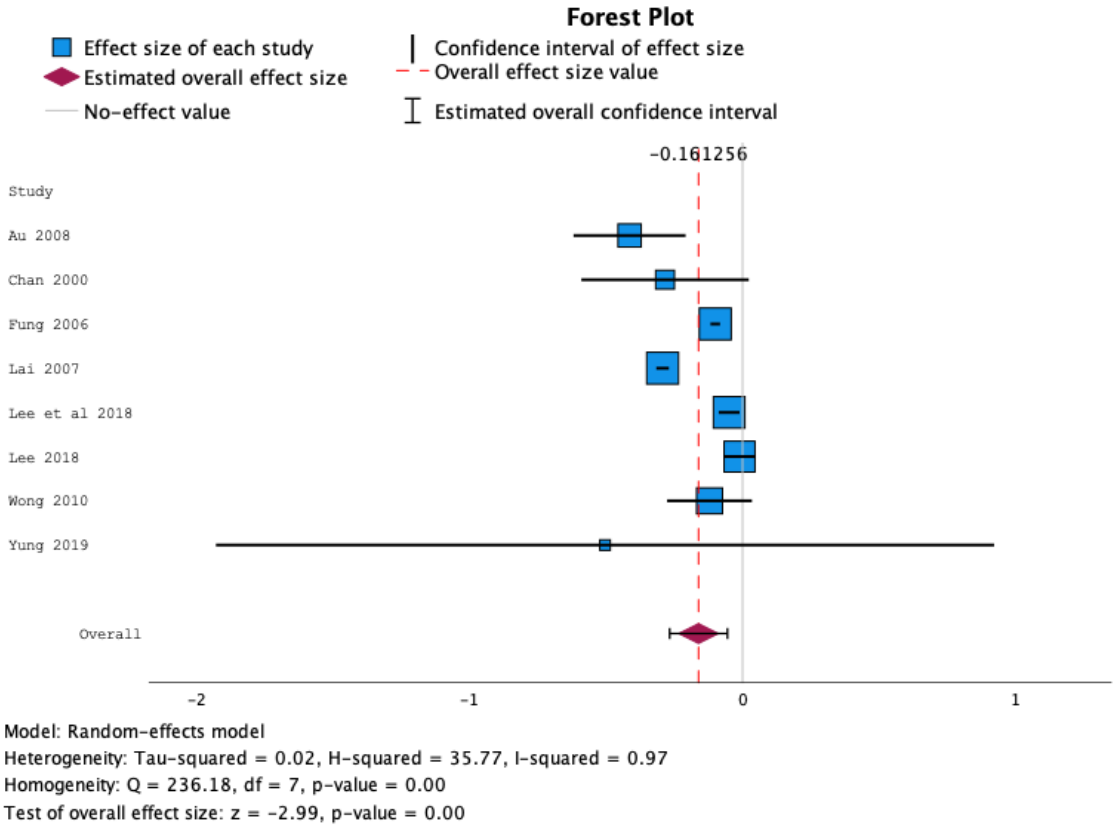


Figure 14. Forest Plot showing all eligible studies

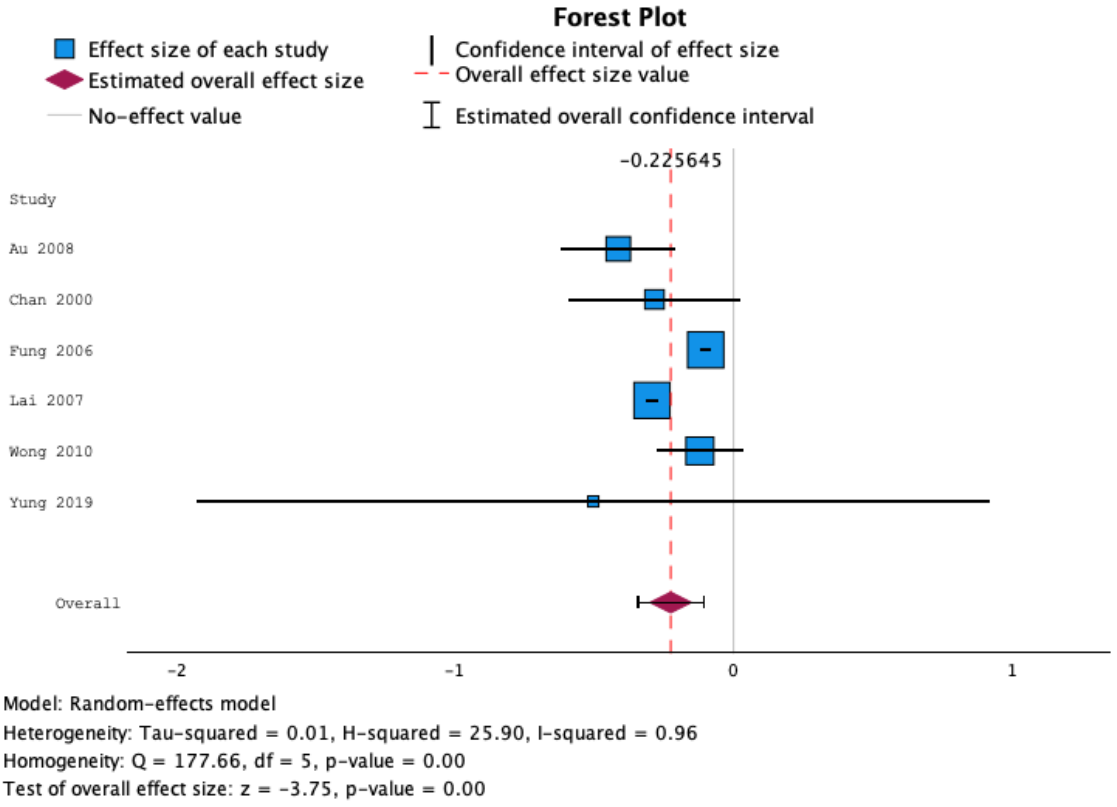


Figure 15. Forest plot showing only studies on general L2 motivation

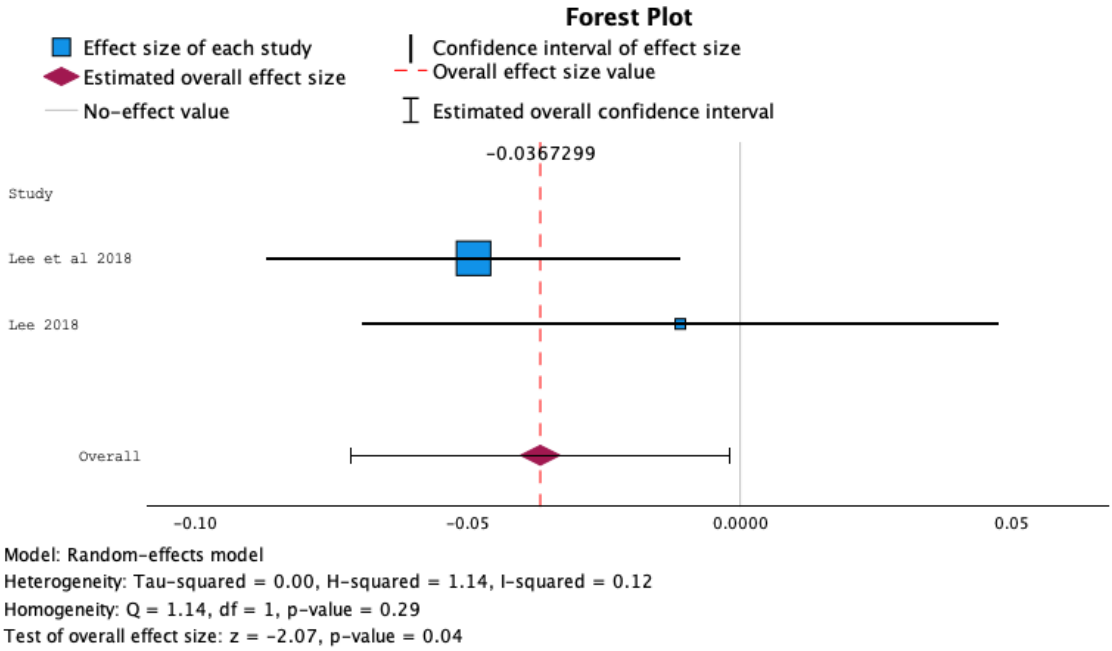


Figure 16. Forest plot showing only studies on specific ESL skills

Table 5. Summary of outcome measures and findings in eligible studies

All values in this table are computed based on available data, unless otherwise specified with a #.

Reference	Outcome measure(s) of ESL motivation	Measurement tools	Male mean (SD)	Female mean (SD)	Cohen's *d*	*p*-value	SEM
Au (2008)	Students' beliefs on motivations and expectations.	4-point Likert scale questionnaire developed based on Horwitz's (1988) Beliefs about Language Learning Inventory (BALLI) survey. 3 items were used to measure students' motivation.	3.233 (0.708)	3.511 (0.602)	-0.414	0.184 (*df*=39)	0.104
Chan (2000)	(i) Intrinsic motivation; (ii) extrinsic motivation.	5-point Likert scale questionnaire adopted from Kruidenier and Clements (1986) and Belmechri and Hummel (1998). There are 12 items on motivation.	3.018 (1.410)	3.403 (1.258)	-0.284	0.216 (*df*=73)	0.157
Fung (2006)	(i) Motivation intensity; (ii) desire to learn; (iii) attitude toward learning English;	A 5-point Likert scale questionnaire with 47 items was used. The questionnaire modified Gardner's (1985b) AMTB, and measured "learning orientation" (10 items), "attitude toward learning English" (10 items), "attitude toward bilingualism"	2.996 (1.329)	3.126 (1.264)	-0.100	0.000*** (*df*=20984)	0.009

Reference	Outcome measure(s) of ESL motivation	Measurement tools	Male mean (SD)	Female mean (SD)	Cohen's *d*	*p*-value	SEM
	(iv) attitude toward bilingualism; (v) learning orientation.	(12 items), "motivation intensity" (7 items), and "desire to learn" (8 items).					
Lai (2007)	(i) Integrative orientation toward English; (ii) instrumental orientation toward English.	A self-developed 4-point Likert scale questionnaire with 24 items was used. Among the items, 8 measured learning orientation toward English.	3.215 (0.501)	3.350 (0.397)	-0.293	0.000*** (*df*=1046)	0.011
I. Lee, Yu, and Liu (2018)	(i) Importance of writing; (ii) ideal L2 self; (iii) ought-to self;	A 5-point Likert scale questionnaire with 40 items was used. The questionnaire measured: (i) importance of writing (language level) (items 1-8); (ii) ideal L2 self (items 9-11); (iii) ought-to L2 self	-0.11 (0.77)[#]	0.00 (0.67)[#]	-0.049	0.004** (*df*=1393)	0.019

Reference	Outcome measure(s) of ESL motivation	Measurement tools	Male mean (SD)	Female mean (SD)	Cohen's *d*	*p*-value	SEM
	(iv) interest in writing; (v) self-regulation; (vi) writing efficacy; (vii) classroom writing context.	(items 12-15); (iv) interest in writing (learner level) (items 16-17); (v) self-regulation (learner level) (items 18-20); (vi) writing efficacy (learner level) (items 21-30); and (vii) classroom writing context (learning situation level/L2 language experience) (items 31-40).					
K. W. Lee (2018)	(i) Autonomous English reading motivation; (ii) Controlled English reading motivation.	A 5-point Likert scale questionnaire, modified from the Self-Regulation Questionnaire of Self-Determination Theory questionnaire, was used. There are 12 items measuring autonomous English reading motivation, and 11 items measuring controlled English reading motivation.	2.765	2.775	-0.011	Unknown due to missing SD.	0.030

Reference	Outcome measure(s) of ESL motivation	Measurement tools	Male mean (SD)	Female mean (SD)	Cohen's *d*	*p*-value	SEM
Wong (2010)	(i) Intrinsic motivation; (ii) extrinsic motivation.	A self-developed 5-point Likert scale questionnaire with 16 items was used. 8 items measured intrinsic motivation (i.e. items 3, 4, 9, 10, 11, 12, 13, 15) and the other 8 measured extrinsic motivation (i.e., items 1, 2, 5, 6, 7, 8, 14, 16).	4.150 (1.337)	4.270 (0.500)	-0.121	0.468 (*df*=154)	0.079
Yung (2019)	(i) Intrinsic motivation; (ii) extrinsic motivation.	A yes/no questionnaire was created to elicit participants' attitudes towards learning English. There are 14 yes/no items, and 1 open-ended question for additional answers, but only one participant attempted the item.	5.889 (2.667)	7.444 (3.432)	-0.504	0.299 (*df*=16)	0.728

4.4. Moderators of L2 Motivation

A narrative synthesis was conducted to explore the moderators of findings. Although a metaregression was planned in the protocol, the small number of eligible studies and missing data prevented a statistical analysis of moderators in SPSS. As advised by Thomson and Campbell (2020), when there is a lack of data, a narrative synthesis can be used to organize, describe, explore, and interpret explanations and moderators of findings. A narrative synthesis adopts a textual approach and manipulates statistical data (if available) to describe the effects and variations in eligible studies.

From the eligible studies, four factors moderating the gender differences in L2 motivation were identified. They are: (i) age, (ii) medium of instruction, (iii) school banding, and (iv) L2 proficiency. This section first outlines the extent of missing data, followed by an exploration of each of the moderators.

4.4.1. Missing Data

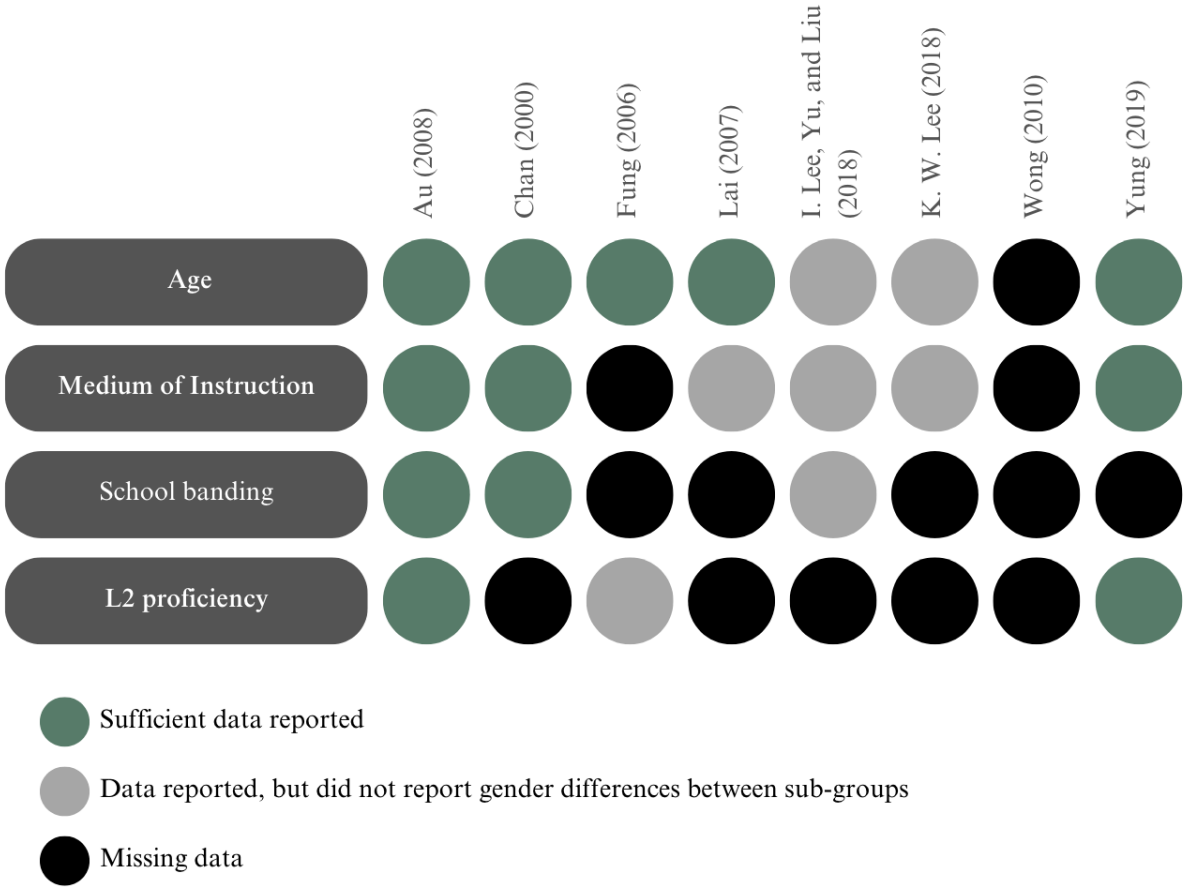


Figure 17. Missing data on moderators across studies

Figure 17 shows missing data on moderators across studies. Original authors of all studies were contacted to obtain the missing data, but to no avail. According to the *Cochrane Handbook for Systematic Reviews of Interventions*, missing data should not be a reason to exclude a study from a systematic review (J. P. T. Higgins et al. 2023). Hence, none of the eligible studies were excluded for missing data, and each moderator is considered separately for the narrative synthesis.

For each moderator, only studies with green dots are included in the analysis since there is sufficient data to analyse the moderators' impact on gender differences in L2 motivation. Studies with black and grey dots are excluded as the former did not report any relevant data, while the latter did not report crossover data on the moderator and gender differences in L2 motivation. Therefore, the synthesis includes five studies for the analysis of age effect, three for the medium of instruction (MoI), two for school banding, and two for L2 proficiency.

Additionally, to conduct *t*-tests for moderators, the standard deviation of difference of means is required. It is computed with the male and female sample sizes and SDs, using the formula below (Moore, McCabe, and Craig 2014):

$$SD_{\text{diff}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

4.4.2. Moderating Effect of Participants' Age

Figure 18 shows a bubble plot of participants' age by the observed gender differences (i.e., effect size) in L2 motivation. Across the five studies, seven clusters of participants were identified, as Chan (2000) included participants from three age groups. Due to missing data, no Secondary 5 students were included in this analysis. In all included studies, negative effect sizes were observed, indicating that female students were more motivated than males in learning ESL. A one-way ANOVA, weighted by sample size, revealed a significant age effect ($t(1)$=-3.178, B=-0.149, F=10.100, p=0.025). This suggests that the older the participants, the wider the observed gender differences in L2 motivation, with females being more motivated than their male counterparts.

Figure 19 shows a scatterplot of male and female participants' L2 motivation by age. As clearly observable from the graph, both genders exhibit a decline in L2 motivation as they age, with males declining at a faster rate than females. From Table 6, the gender difference is insignificant in Secondary 1 (k=1, $t(23)$=0.338, p=0.739) and Secondary 2 (k=1, $t(64)$=0.659,

p=0.512), but becomes significant in Secondary 3 (*k*=2, *t*(21009)=7.321, *p*=0.000) and Secondary 4 (*k*=1, *t*(1046)=4.861, *p*=0.000). There is no data on Secondary 5 students. The gender difference then becomes insignificant again in Secondary 6 (*k*=1, *t*(16)=1.074, *p*=0.299). Nonetheless, upon a statistical analysis using one-way ANOVA, no significant change in L2 motivation was found in both genders combined over time (*F*=6.157, *t*=2.481, *p*=0.056), or separately in males (*F*=6.496, *t*=2.549, *p*=0.051) and females (*F*=5.858, *t*=2.420, *p*=0.060).

Table 6. Significance of gender difference by age group

		Male			Female					
Age	*k*	mean	SD	*n*	mean	SD	*n*	*df*	*t*	*p*
S1	1	0.687	0.243	14	0.720	0.245	11	0.338	23	0.739
S2	1	0.771	0.203	28	0.804	0.205	38	0.659	64	0.512
S3	2	0.599	0.266	11071	0.625	0.253	9940	7.321	21009	0.000
S4	1	0.804	0.125	555	0.838	0.099	493	4.861	1046	0.000
S5	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
S6	1	0.421	0.190	9	0.532	0.245	9	1.074	16	0.299

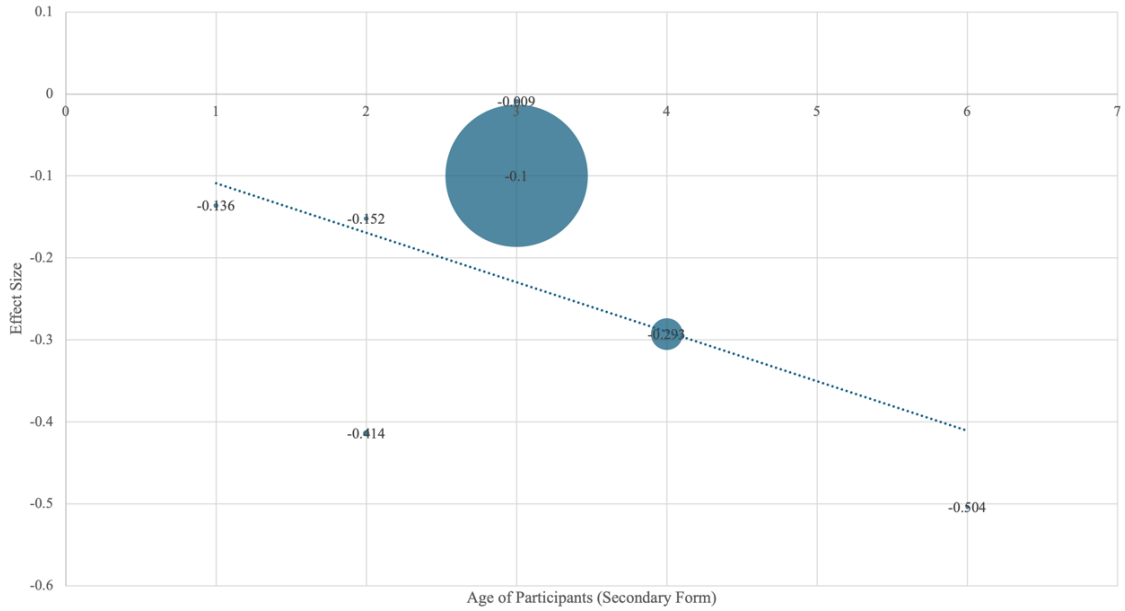


Figure 18. Bubble plot showing effect size by the age of participants

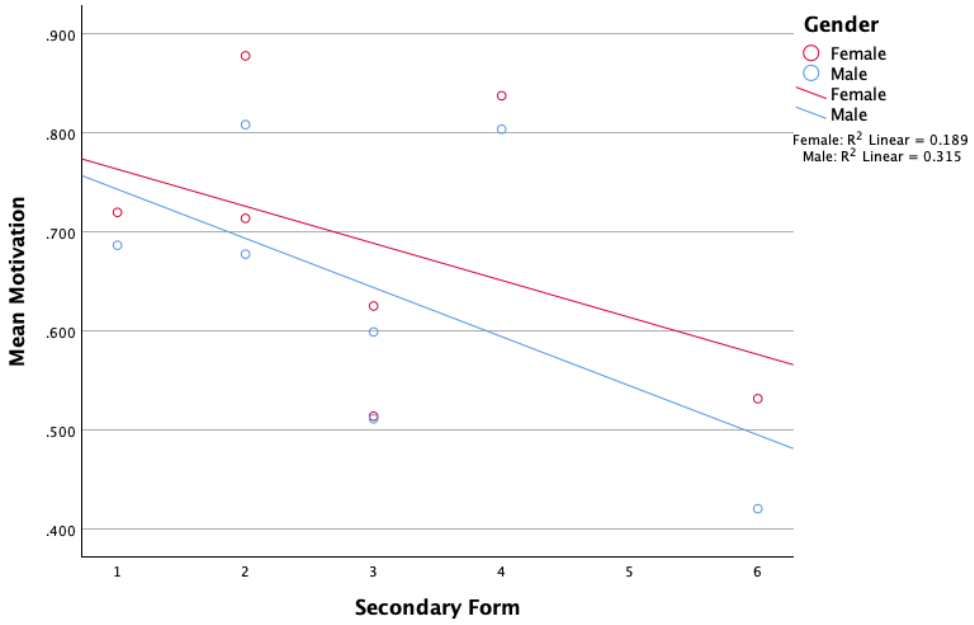


Figure 19. Scatterplot showing motivation of participants by gender and age

4.4.3. Moderating Effect of the Medium of Instruction (MoI)

Figure 20 shows a bubble plot of effect size by the medium of instruction at participants' schools. An independent samples *t*-test revealed that the gender differences in L2 motivation did not significantly differ by MoI ($t(132)=0.486$, $p=0.628$). This suggests that the observed gender differences in L2 motivation are not significantly different in CMI schools and EMI schools. Additionally, in general, students' L2 motivation did not differ by the schools' MoI ($t(132)=0.634$, $p=0.527$).

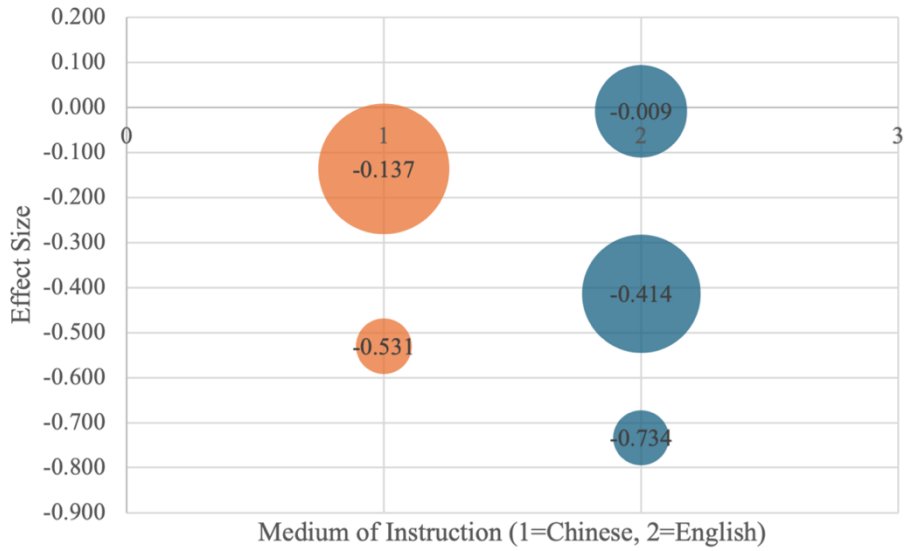


Figure 20. Bubble plot showing effect size by the medium of instruction

4.4.4. Moderating Effect of School Banding

Only two studies reported sufficient data on school banding. One of the studies, Chan (2000), reported that the participants came from a Band 2-3 school. Hence, participants from that study were coded as Band 2.5 students. An independent samples *t*-test revealed no significant differences in the observed gender differences in Band 1 and Band 2.5 schools ($t(114)=0.693$, $p=0.490$).

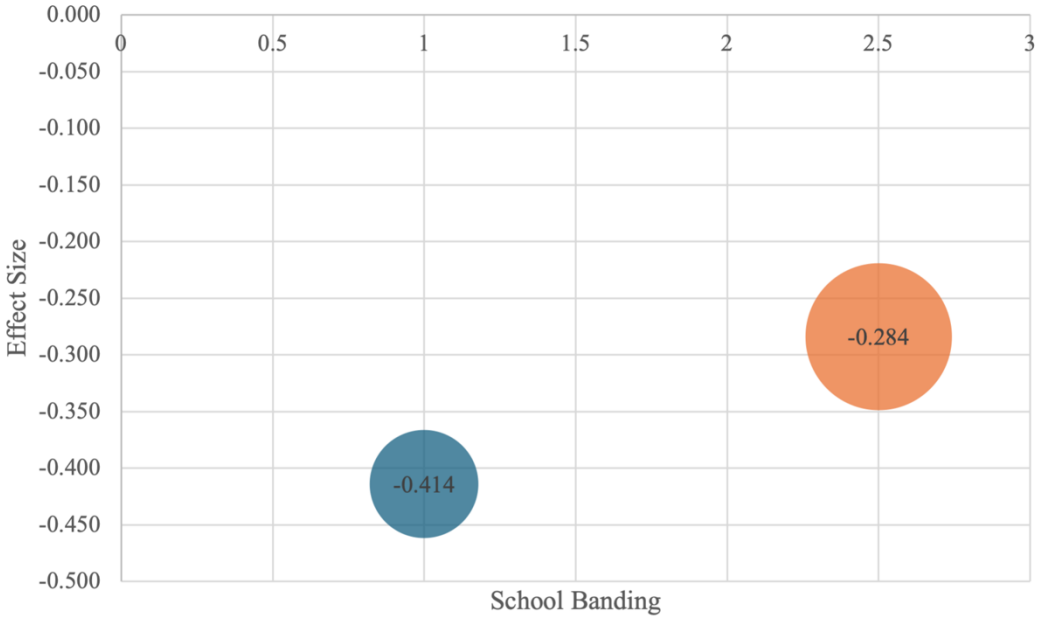


Figure 21. Bubble plot showing effect size by school banding

4.4.5. Moderating Effect of L2 Proficiency

Only two studies reported sufficient data on the participants' L2 proficiency. When coding for L2 proficiency, both studies used ordinal scales instead of ratio scales. In Au (2008), participants were asked to self-rate their most recent intra-school examination result; the rating categories were "fail," "pass," "satisfactory," "credit," and "distinction." In Yung (2019), students' L2 proficiencies were rated based on their performances in the public examination, i.e., the Hong Kong Diploma of Secondary Education (HKDSE). Students' proficiencies are ranked by seven levels, from lowest to highest: 1, 2, 3, 4, 5, 5*, and 5**. For both of the studies, the ordinal scores were first converted into numerical values (i.e., 1 to 5 and 1 to 7 respectively). Since the studies used different scales, the scores were then

standardized to a 0 to 1 scale. The gender differences in L2 proficiency were then calculated, as shown in Figure 22.

From the two studies, no meaningful conclusion could be drawn as there are no significant gender differences in either L2 proficiency (Au (2008): $t(39)=1.818$, $p=0.0767$; Yung (2019): $t(16)=1.860$, $p=0.0814$) or L2 motivation (Au (2008): $t(39)=1.351$, $p=0.184$; Yung (2019): $t(16)=1.073$, $p=0.299$). Hence, it is unknown whether a difference in L2 proficiency could have predicted a difference in L2 motivation.

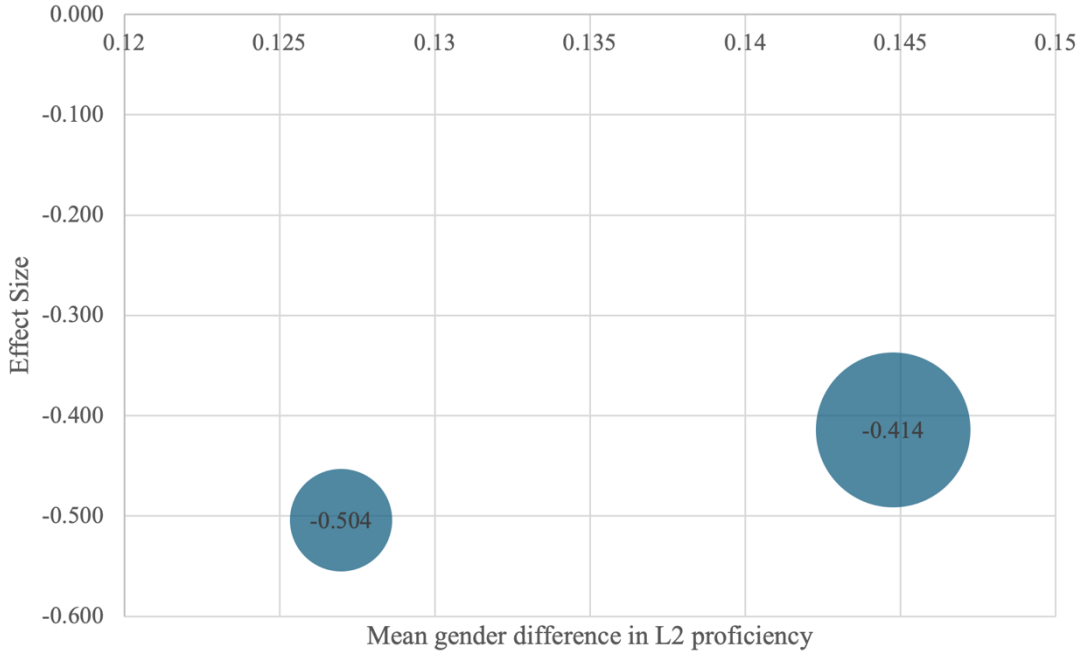


Figure 22. Bubble plot showing effect size by mean gender difference in L2 proficiency

4.5. Publication Bias

Publication bias refers to the phenomenon where studies with insignificant results are less likely to be published (Chalmers, Brown, and Koryakina 2023). To check for publication bias, an Egger's regression was carried out. The eligible studies were coded for their publication type (i.e., journal article, thesis) as well as the gender of the first author. The results showed no publication bias in terms of both publication type ($t=0.438$, $p=0.684$, SEM=0.133) and the first author's gender ($t=-1.037$, $p=0.358$, SEM=0.160). Figure 23 and Figure 24 show symmetrical funnel plots, indicating an absence of publication bias.

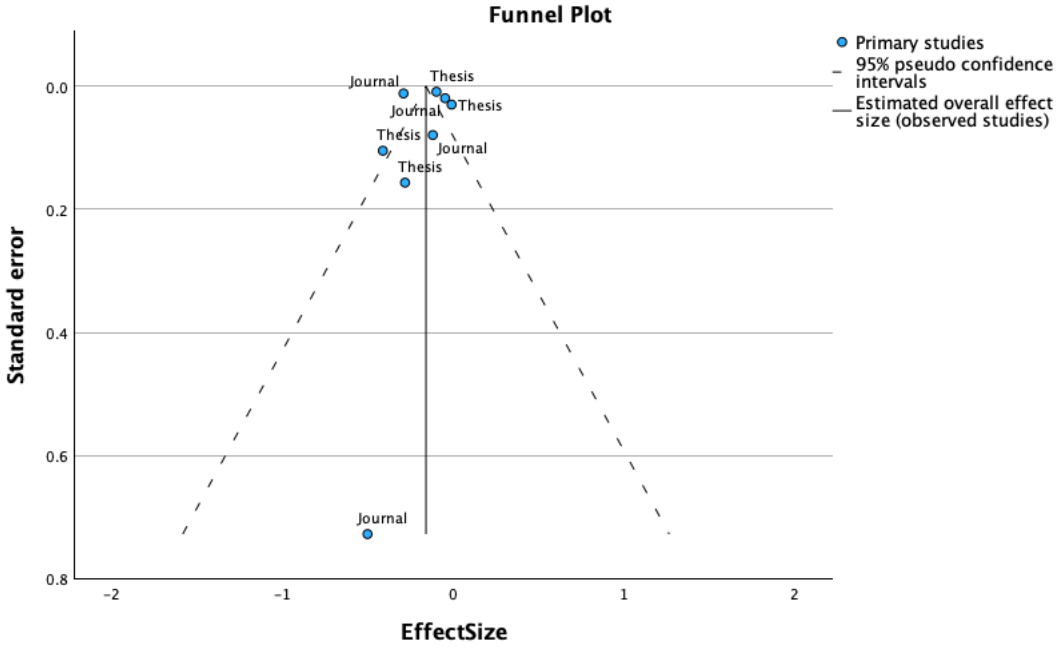


Figure 23. Funnel plot by publication type

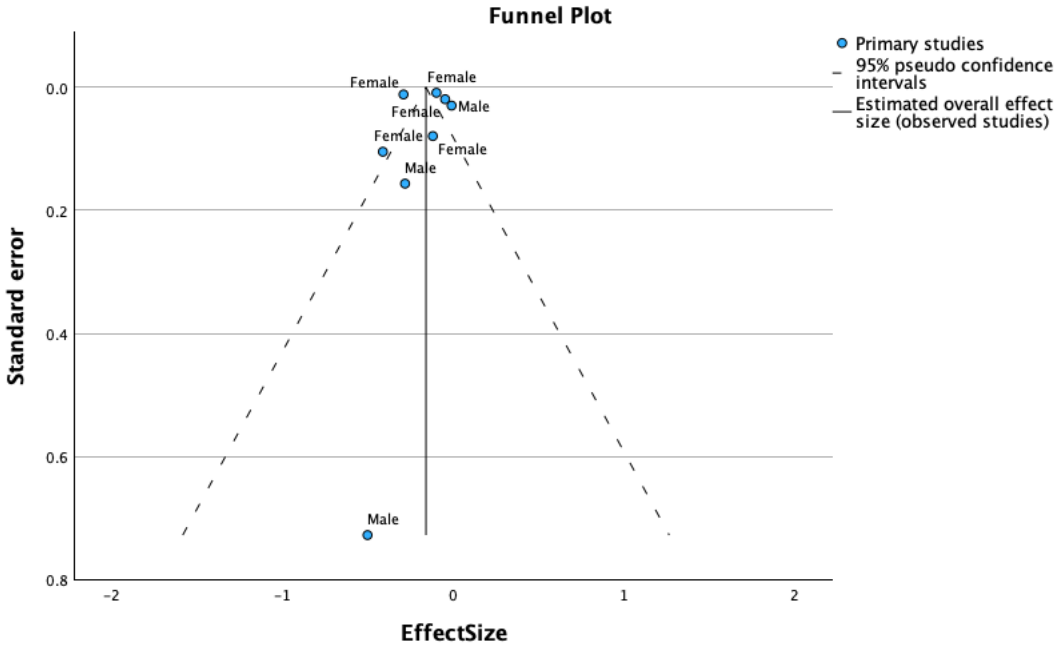


Figure 24. Funnel plot by the first author's gender

4.6. Risk of Bias

Table 7 summarizes the risk of bias of individual studies. The higher the QuADS score, the lower of the risk of bias. In general, most studies had a moderate risk of bias, scoring between 60% and 80% (Harrison et al. 2021). The QuADS scores ranged from 22 (56.4%) to 33 (84.6%) out of 39. Of all the criteria, the studies performed best at criterion 3 (average score=3/3), as all studies provided clear descriptions of the research setting and

target populations. Conversely, the studies scored the lowest in criterion 12 (average score=1.375/3), indicating that there is little use of pilot studies, project advisory groups, or other forms of stakeholder input. For full scoring criteria, see Appendix B.

Interestingly, a regression analysis at 95% confidence interval revealed a significant and strong correlation between the QuADS score and the observed effect size (t=3.555, p=0.012, r=0.823). From Figure 25, the higher the QuADS score (i.e., the lower the risk of bias), the more likely the effect size tends to zero. In other words, in a well-designed study free of bias, it is more likely to observe no gender differences in L2 motivation among Hong Kong secondary school students.

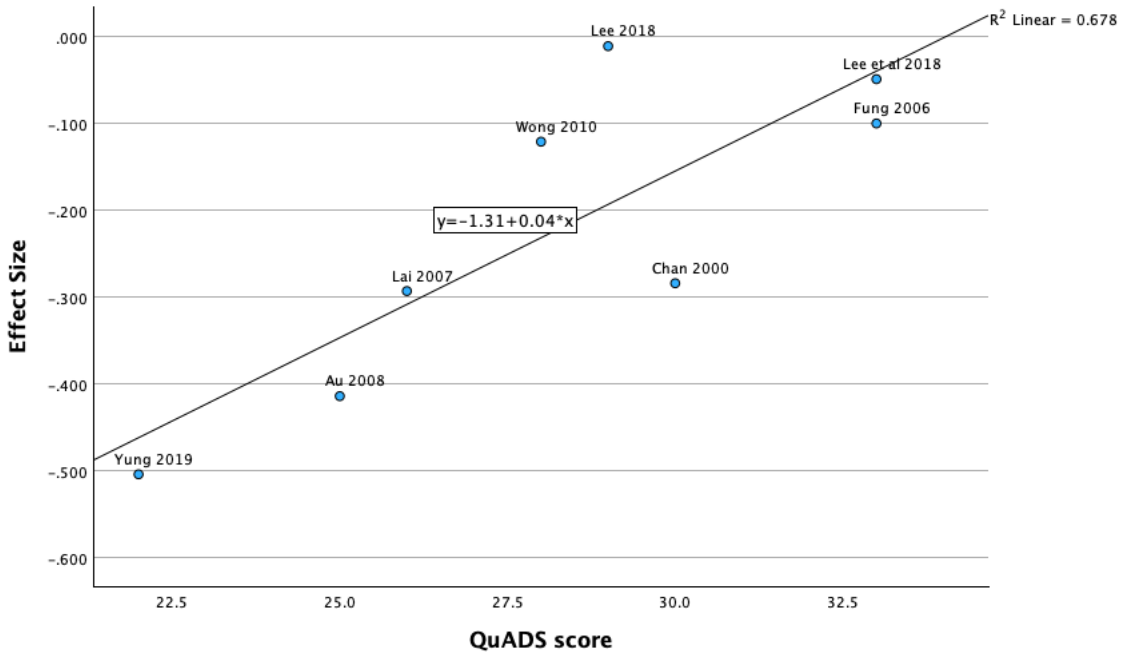


Figure 25. Scatter plot of effect size by QuADS score

4.7. Cumulative Confidence across Studies

Table 8 summarizes the cumulative confidence across studies using the GRADE assessment tool. Overall, the body of literature was given a score of 0, indicating that the quality of cumulative evidence was “very weak”. This is due to the inconsistency of findings, risk of bias across studies, indirectness, and imprecision of individual studies. For full rubrics, see Appendix C.

Table 7. QuADS rating of individual studies

	Au (2008)	**Chan (2000)**	**Fung (2006)**	**Lai (2007)**	**I. Lee, Yu, and Liu (2018)**	**K. W. Lee (2018)**	**Wong (2010)**	**Yung (2019)**	**Average**
1. Theoretical or conceptual underpinning to the research	2	3	3	2	3	3	3	1	2.5
2. Statement of research aim/s	3	3	3	2	3	3	3	2	2.75
3. Clear description of research setting and target population	3	3	3	3	3	3	3	3	3
4. The study design is appropriate to address the stated research aim/s	3	2	2	3	2	2	3	3	2.5
5. Appropriate sampling to address the research aim/s	0	2	3	2	2	1	1	3	1.75
6. Rationale for choice of data collection tool/s	2	2	3	2	3	2	3	1	2.25
7. The format and content of data collection tool is appropriate to address the stated research aim/s	2	2	3	2	3	3	2	1	2.25

	Au (2008)	Chan (2000)	Fung (2006)	Lai (2007)	I. Lee, Yu, and Liu (2018)	K. W. Lee (2018)	Wong (2010)	Yung (2019)	Average
8. Description of data collection procedure	3	3	2	1	1	3	3	2	2.25
9. Recruitment data provided	2	3	2	1	1	1	1	3	1.75
10. Justification for analytic method selected	0	3	3	3	3	3	1	1	2.125
11. The method of analysis was appropriate to answer the research aim/s	1	2	3	2	3	2	2	1	2
12. Evidence that the research stakeholders have been considered in research design or conduct	3	2	0	0	3	0	2	1	1.375
13. Strengths and limitations critically discussed	1	0	3	3	3	3	1	0	1.75
Total score (out of 39)	25	30	33	26	33	29	28	22	
Risk of bias	moderate	moderate	low	moderate	low	moderate	moderate	high	

Table 8. Cumulative confidence across studies (GRADE)

Criteria	Score awarded	Description/Reason
Study Design	+4	A body of literature begins with an assumption of "high" certainty (J. P. T. Higgins et al. 2023).
Risk of bias	-1	Most information is from studies at moderate risk of bias.
Inconsistency	-2	The Q-test shows a low *p*-value ($Q(7)=236.178$, $p<0.001$) and a substantial I^2 (97.2%).
Indirectness	-1	Some differences exist between the populations studied across studies such that the magnitude of estimated effect might differ moderately. Some differences exist between the outcome measures across studies. Head-to-head comparisons between two or more interventions are somewhat feasible.
Imprecision	-1	Modest sample size with moderate confidence interval.
Publication bias	0	Egger's regression shows little to no publication biases.
Large effect	0	There is an absence of large effect (i.e., $d \geq 0.8$) across studies.
Dose-response gradient	0	This criterion is inapplicable as there were no interventions involved.
All plausible confounding	1	It is likely that unmeasured confounding variables (e.g., age) have lowered the estimated effects.
Total score	0	(very weak)

5. DISCUSSION

The aim of this meta-analysis was to identify potential gender differences in secondary school students' motivation to learn English as a second language in Hong Kong, and to explore the effect of moderators. Despite the rise of motivational research in recent years (Boo, Dörnyei, and Ryan 2015), this study identified a mere eight studies that explored motivational differences among Hong Kong secondary school students and listed gender as one of the variables. Furthermore, given the low cumulative confidence across the eligible literature, one must proceed with caution before making any inferences with the reported outcomes. Nonetheless, the body of literature offers novel insights into possible gender effects on L2 motivation in this educational context, and suggests paths for more robust motivational research in the future. Section 5.1 first interprets the findings, and discusses the limitations and implications of findings in relation to previous studies. Section 5.2 then outlines some pedagogical implications, followed by the limitations in Section 5.3.

5.1. Interpretation and Implications of Results

5.1.1. RQ1: Extent and Nature of Empirical Research

Using pre-registered search strings and eligibility criteria, this meta-analysis identified eight empirical studies covering 24,660 participants from all secondary school age groups (i.e., Secondary 1 to Secondary 6). Of the eight eligible studies, six studied general L2 motivation, while the other two studied English reading and writing motivations. There was no observable change in the research trend, whether in the number of studies or in methodological types (i.e., quantitative and mixed-methods). To allow for a meta-analysis, all eligible studies employed quantitative methods to generate statistical outcomes; three out of eight studies also applied qualitative measures to enhance their studies' depths. Most of the studies explicitly stated the theoretical frameworks on which their measures were based, including the Socio-Educational Model (S-EM), Self-Determination Theory (SDT), and the L2 Motivational Self-System (L2MSS). The measures include intrinsic/extrinsic motivation, integrative/instrumental motivation, attitudes towards English and bilingualism, motivational intensity, desire to learn English, Ideal L2 self, Ought-to Self, and measures of reading and writing motivation. However, the measures were often irrelevant to the motivational constructs concerned, and the instruments were seldom validated. This will be explained in the Section 5.1.4.

5.1.2. RQ2: Gender and Hong Kong Secondary School Students' L2 Motivation

Using an array of measures for L2 motivation, all eight studies found female students to be more motivated than male students in learning ESL, albeit to varying degrees. For studies on general L2 motivation, the combined data found a significant but weak effect size favouring females, suggesting that female participants are slightly more motivated than their male counterparts. Nonetheless, when the studies only concern specific ESL skills (e.g., reading motivation, writing motivation), the gender difference is significant but so small that it is negligible. Overall, combining the studies on general motivation and specific ESL skills, the eight studies have found a significant but negligible gender difference. This indicates that there are indeed gender differences in L2 motivation among Hong Kong secondary schools students, but the difference is so minimal that it is not easily observed in daily contexts.

The findings of this meta-analysis echoes findings from other meta-analyses (e.g., Qi 2021; Yousefi and Mahmoodi 2022) and some individual studies (e.g., Dörnyei and Clement 2002; Dörnyei, Csizér, and Németh 2006), all of which found significant gender differences in L2 motivation despite differences in measures. One contribution of the current review is the finding that the effect size of gender largely depends on the study focus: whereas studies on general L2 motivation observed a significant and small effect size, studies focusing on specific ESL skills, such as reading and writing motivations, found an effect size so small that it was negligible. This study also contributed to the field of motivation by focusing on the unique context of Hong Kong, a bilingual city where British/Western and Chinese/Eastern cultures intersect. The study anticipated that the unique gender norms in Hong Kong, informed by both Eastern and Western values, might produce different patterns in L2 motivation as observed by the rest of the world. However, the findings suggested that the case of Hong Kong largely coincided with the global trend: female students were significantly more motivated than boys in L2 learning, although the difference was very small.

A possible explanation for the gender difference lies in the gender ideology surrounding language learning. Extensive literature has established the role of gender ideologies in influencing L2 learning practices, such as learners' interest (Chaffee et al. 2020), self-efficacy (Kutuk et al. 2022), and goals and trajectory (Brutt-Griffler and Kim 2016; 2018). Previous studies have observed that language subjects are still considered "feminine" in Hong Kong and across the world (Clark and Trafford 1995; Lai 2007; Zhang 2011). Given the stereotype, a large-scale study by Chaffee et al. (2020) reported that the femininity associated with language subjects might pose a "masculinity threat" to Canadian

undergraduates with "traditional masculinity ideologies." This led them to report less interest in language learning and language majors, although it did not have a significant effect on career interest. A qualitative study conducted in Hong Kong, focusing on two female undergraduates' experiences in learning English, further complicates the issue by suggesting that one's motivation is not simply shaped by external, social ideologies, but also by their internal negotiation with these ideologies. Sung (2022) reported two distinct cases of how female students experienced such gender ideologies: One female undergraduate participant, who accepted traditional gendered expectations for a woman's behaviours, aspired to become an English teacher and found herself being more agentive in developing native-like fluency in English. Conversely, another female participant, who was more willing to challenge traditional gender norms, had different motivations to learn English, and used English learning activities (e.g., public speaking) as a masculine capital to construct her identity.

Indeed, in studies involving gender differences in L2 learning, it is important not to essentialize gender qualities and behaviours. While there may be distinct social expectations for each gender, the relationship between social expectations and social behaviour are not linear (i.e., women are expected to perform better at languages, so all women will be more motivated to learn English). As Sung (2022) noted, one's engagement in L2 learning practices are shaped not only by social gender ideologies, but also by how they understand, experience, and respond to such ideologies.

Lastly, it is worthwhile to discuss why this meta-analysis found such a small gender difference. While the existing literature has established that language subjects are still perceived as "feminine" in Hong Kong, this study speculates that some languages might be considered more "feminine" than others, and English might be gender-neutral in Hong Kong. This hypothesis stems from Dörnyei, Csizér, and Németh's (2006) longitudinal study: They surveyed Hungarian students and found that female participants preferred French and Italian, male students preferred German and Russian, but English was found to be gender-neutral. Nonetheless, regardless of their language preferences and their actual L2, girls showed more commitment than boys in language learning, as measured by Intended Effort. The authors argued that gender-neutrality of English might come from its increasingly prominent status as a global language, making it a language that both genders must learn. This argument is supported by Lai's (2007) qualitative analysis conducted in Hong Kong: participants of the study found English important for all students, as it is a compulsory subject in the secondary school curriculum, and a core requirement for university entry. As such, the necessity of

English Language to academic success in Hong Kong might have been the reason for the narrow gender gap. More mixed-methods and qualitative studies in the future would be required to evaluate this claim.

5.1.3. RQ3: Moderators of Gender Differences in L2 Motivation

This meta-analysis identified four moderators of gender differences in L2 motivation: age, medium of instruction (MoI), school banding, and L2 proficiency. The moderators were identified based on moderating factors listed in the eight studies. This study found a significant moderating effect of age, but no significant effect of MoI, school banding, and L2 proficiency. It is speculated that the insignificant results were largely due to missing data. As most studies did not report the gender differences within each moderating sub-group (e.g., gender gaps in CMI and EMI schools, respectively), it was difficult to deduce how the moderators affected the gender difference in L2 motivation. Of the eight eligible studies, only three reported sufficient data on MoI, three on school banding, and two on L2 proficiency. While bubble plots could still be created to visualize the moderating effect, convincing conclusions could not be derived from the limited data. Nonetheless, the limited studies served as case studies, and provided a window into the confounding factors that could impact gender differences in L2 motivation.

5.1.3.1. Moderating Effect of L2 Proficiency and School Banding

The current meta-analysis found insufficient data on the moderating effects of L2 proficiency and school banding, as only two studies were included for each moderator. This points to a need for empirical studies exploring these two factors' impact on the gender gap in L2 motivation, particularly in the context of Hong Kong.

Regarding L2 proficiency, existing meta-analyses have been equivocal about the correlation between L2 motivation and L2 proficiency (measured by L2 grades). For instance, while Masgoret and Gardner (2003) found weak correlations between L2 grades and constructs of the S-EM (r(attitudes towards learning situation)=0.24, r(integrativeness)=0.24, r(motivation)=0.37, r=(integrative orientation)=0.20), Al-Hoorie (2018) found no significant relationship between L2 grades and constructs of the L2MSS (r(IS)=0.103, r(OTS)=-0.048, r(L2LE)=0.111). The different results might have arisen from the different motivational theories applied, or the variety of instruments used to measure one's L2 proficiency.

Moreover, regarding school banding, this meta-analysis found insufficient data to draw a convincing conclusion. In truth, few studies conducted on Hong Kong secondary schools have included school banding as a variable. One of the few exceptions is I. Lee, Yu, and Liu (2018), which found significant differences among students from different bandings in their L2 writing motivation. However, as the study focused on writing motivation and did not report the gender differences within each banding, it cannot be deduced whether school banding influences the extent of gender differences in L2 motivation.

5.1.3.2. Moderating Effect of the Medium of Instruction

Furthermore, the current study found that (i) students' L2 motivation did not significantly differ by MoI, and (ii) the gender differences in L2 motivation are not significantly different in CMI and EMI schools. This finding contradicts the findings in Lo and Lo's (2014) meta-analysis on Hong Kong secondary schools, which found EMI students to exhibit significantly higher learning motivation than CMI students (mean d=0.02, p<0.05) and higher self-concept in English (mean d=0.28, p<0.01). While Lo and Lo's meta-analysis studied general learning motivation instead of L2 motivation specifically, the stark contrast between CMI and EMI students across all measures (e.g., self-concept, motivation, strategy, interests) suggest that motivational differences might also exist in L2 learning. The insignificant results in the current review might be a result of the small sample size (k=3, n=134), which calls for more empirical studies in the future.

All in all, there is limited data on the moderating effects of L2 proficiency, school banding, and MoI on students' gender differences in L2 motivation. It is recommended that empirical studies in the future report the male and female means separately for each moderating factor, such that the gender differences in L2 motivation could be investigated.

5.1.3.3. Moderating Effect of Age

Of all the factors studied, only age showed a significant moderating effect. Examining seven clusters of participants from five studies, descriptive statistics indicated that motivation declined for both groups, though neither within-group difference was found to be significant. A one-way ANOVA revealed that the gender gap in L2 motivation significantly widened with age, with male students' motivation declining at a faster rate than females. Notably, a significant gender difference in L2 motivation emerged in S3, and the effect carried on to S4, until the gender difference eventually became insignificant again in S6. The gender difference

in S5 was not investigated due to missing data. The findings suggest that all students' L2 motivation dropped very little throughout six years of secondary education, but the gender difference in L2 motivation would gradually widen to a significant level.

Unlike previous studies, the current meta-analysis found no statistical support for the decrease in L2 motivation across age. A significant decrease in L2 motivation was initially anticipated in this review, as previous literature suggested so. For example, Gardner and Smythe (1975) studied a group of Canadian students over two years, and found that their motivation to learn French as an L2 dropped from Grade 7 to Grade 10. This finding was corroborated by Bećirović and Hurić–Bećirović (2017) and Ghenghesh (2010). However, while there is an observed decline in the mean scores for motivation over the years (Figure 19), this meta-analysis found no significant change in L2 motivation across ages in either of the sexes or combined data. This is possibly due to the small sample size, as only seven clusters of participants (*n*= 22,168) were identified for this moderator, and data for S5 students were missing.

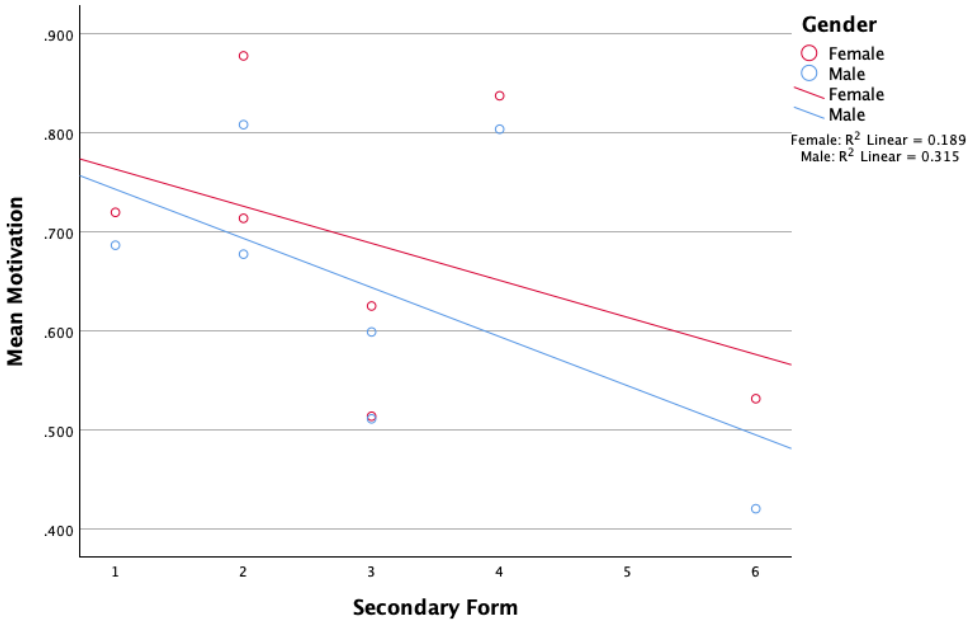


Figure 19. Scatterplot showing motivation of participants by gender and age

A more prominent finding from this review is that the gender difference in L2 motivation significantly widens with age. This echoes Maccoby and Jacklin's (1974) groundbreaking observation that boys' and girls' language proficiencies, measured by verbal skills, begin to differentiate during adolescence. The fact that L2 motivation and L2 abilities both diverge during adolescence points to a possible correlational/causal relationship between

the two. A more recent study by MacIntyre et al. (2003) also supports this review's findings. The study measured Canadian high school students' L2 motivation by their willingness to communicate, and found that while there were no significant gender differences in Grade 7 and 8, girls' willingness to communicate became significantly higher than boys' in Grade 9. This mirrored the findings from this review: no significant gender differences in S1 (Grade 7) and S2 (Grade 8) students were found, but the gender difference became significant in S3 (Grade 9) and S4 (Grade 10). In S6, the L2 motivation of both sexes continued to drop, but the gender difference became insignificant again.

Since the qualitative findings from individual studies were not synthesized in this review, one cannot definitively conclude why the gender gap emerged in S3, but not sooner or later. The author suspects, however, that the gender gap emerges among Hong Kong secondary school students as a result of career-choosing in S3. In Hong Kong, S3 students have to choose their electives for the HKDSE public examination; most students base their choices on the university majors they want to take, or the future careers they want to pursue. Since art subjects such as English are still considered "feminine" in Hong Kong (Lai 2007; Zhang 2011), female students might be motivated to perform better in language subjects, such that they have enough scores to choose the electives they prefer. On the other hand, while male students are still required to study English, they might be inclined to divert their attention to science subjects, such that they could choose a career path in science. Later in S6, the gender gap closes again as both sexes have to study for HKDSE; to enter university, all students have to attain at least Level 3 in English Language, which might motivate all students, regardless of gender, to put time and effort into studying English again. To investigate the cause of divergence in S3, more qualitative and/or mixed-methods studies are required in the future.

Lastly, the findings of eligible studies are limited by their cross-sectional nature, which might have overlooked the fluctuational, dynamic characteristics of motivation. As Ramezanzadeh (2021) and Kormos and Csizér (2008) noted, the learners and the motivational mechanisms (e.g., environmental factors) at play are all subject to longitudinal fluctuations. Taking findings from cross-sectional studies only, this meta-analysis cannot examine motivational changes within the academic year, or investigate the same participants' motivational changes over years. This calls for more longitudinal studies in the future to capture the dynamics of L2 motivation among Hong Kong secondary school students.

5.1.4. RQ4: Reliability of Existing Empirical Evidence

Overall, using the GRADE framework, the cumulative confidence across studies is very weak. This is mostly due to the serious risk of bias, heterogeneity in findings, small sample sizes, and heterogeneity in measures. Regarding individual studies, five had a moderate risk of bias, two had low risks, and one had a high risk of bias. Few studies have provided sufficient recruitment data, justifications for the sample size and analytic methods, consulted stakeholders, and discussed their studies' strengths and limitations. A linear regression at 95% confidence interval suggested that studies of lower risks of bias reported significantly smaller gender differences in L2 motivation. Arguably, this could imply that the larger effect sizes found in weaker studies were caused by poor designs, poor methodology, and poor reporting practices. In a study that is well-designed, the effect size found should tend to zero (i.e., no observed gender difference). This suggests that in the real world, the observed gender difference in L2 motivation would be very small, although secondary school girls in Hong Kong are still significantly more motivated than boys in learning English.

A prominent issue in regard to all eight studies concerns their measures. This problem manifests in three dimensions: (i) mismatch with the motivational model used, (ii) oversimplification/incorrect understanding of motivational models, and (iii) the lack of Confirmatory Factor Analysis (CFA) and piloting.

Firstly, there is a clear mismatch between the measures and the motivational theories adopted (Figure 13). For example, in the L2MSS, motivation is originally conceptualized into three constructs: Ideal L2 Self (IS), Ought-to Self (OTS), and L2 Learning Environment (L2LE). To fully measure motivation with L2MSS, then, would require measures in all three constructs, or at least the two self-constructs (i.e., IS and OTS) since the L2LE is still undertheorized (Dörnyei 2019). However, in one of the studies (i.e., Yung 2019), the author measured motivation by "intrinsic motivation" and "extrinsic motivation," which were not employed in the L2MSS framework. This was also the case for another study using the Socio-Educational Model (S-EM). Wong (2010) measured motivation using an intrinsic/extrinsic motivation scale, rather than attempting to capture constructs stipulated in the S-EM, such as attitudes towards learning English, integrativeness, and motivational intensity. As a result of these mismatches, it was determined that there was no meaningful or valid methodological difference between groups of studies using the S-EM, SDT, and the L2MSS; this was also the reason why the studies were not analysed by motivational theories in this meta-analysis. In fact, the inadequate measures and conflation of theories has been

repeatedly criticized in literature (e.g., Al-Hoorie 2018; Dörnyei 2005), indicating that such methodological issues are not unique to this study, but are identified across the field as well.

OUTCOME MEASURES	SOCIO-EDUCATIONAL MODEL	SELF-DETERMINATION THEORY	L2MSS
MODIFIED AMTB*	✓		
INTEGRATIVE/INSTRUMENTAL ORIENTATION	✓		
INTRINSIC/EXTRINSIC MOTIVATION	✓	✓	✓
CONTROLLED/AUTONOMOUS READING MOTIVATION		✓	
IDEAL SELF, OUGHT-TO SELF, AND OTHER MEASURES OF WRITING MOTIVATION**			✓

*Measures in modified AMTB: (i) integrative/instrumental orientation, (ii) attitudes towards English and bilingualism, (iii) motivational intensity, and (iv) desire to learn.
** Measures of writing motivation include: (i) importance of writing, (ii) interest in writing, (iii) self-regulation, (iv) writing efficacy, and (v) classroom writing context.

Figure 13. Outcome measures by theories applied

Secondly, the outcome measures used in the studies often oversimplified the relevant motivational theory. For example, the S-EM measures L2 motivation through four constructs: attitude towards learning English, integrative orientation, instrumental orientation, and motivation. In practice, however, motivational studies often reduce the model to a integrative/instrumental dichotomy, neglecting all the other factors (Dörnyei 2005). This was observed in one of the studies: Lai (2007) used the S-EM as its theoretical framework but only measured participants' integrative and instrumental orientations. The problem of oversimplification was also found in Chan (2000), which adopted the SDT. Ideally, studies using the SDT would measure the source of motivation (intrinsic/extrinsic), as well as regulation (internal/external). In this study, however, the dimension of regulation was omitted. With the inadequacy in measures, the robustness of some studies are compromised.

Thirdly, the studies generally lack piloting and Confirmatory Factor Analysis (CFA). CFA is used to statistically confirm that the intra-structure of a measured construct is consistent with the researchers' understandings of it (Gallagher and Brown 2013). Since most of the studies included used self-developed surveys, CFA is crucial in validating the questionnaires and instruments. Additionally, piloting and consultations with professional

groups (e.g., supervisors, a group of doctoral students) could also improve the reliability and validity of the instruments. However, CFA and piloting were rare in the included studies which poses questions regarding the tools' validity and reliability in measuring the outcome variables.

All in all, the existing empirical evidence has low reliability, both across studies and within individual studies. This points to a need for better-designed, better-justified, and better-reported research in the field to establish more convincing findings.

5.2. Pedagogical Implications

This study has highlighted a significant gender difference in L2 motivation emerging in Secondary 3, a critical period when students make pivotal academic and career decisions. Given the importance of motivation in influencing students' learning behaviour (Yousefi and Mahmoodi 2022) and its potential impact on L2 achievement (Masgoret and Gardner 2003), it becomes imperative for educators to address this disparity. Hence, the study proposes three pedagogical strategies drawn from the findings:

1. **Focused motivation strategies, especially for boys in Secondary 3:** Descriptive statistics from this study identified a gradual, albeit insignificant, drop in motivation across age, and a significant divergence in L2 motivation among boys and girls in S3. Teachers could consider tailoring their motivational strategies to re-engage students, especially boys, during this critical stage. This could involve incorporating more interactive activities that align with their interests and learning styles, such as the use of gamification (Azzouz Boudadi and Gutiérrez-Colón 2020). Furthermore, teachers could also strive to create a gender-inclusive classroom that encourages both boys and girls to participate equally in L2 learning. This could include using gender-neutral language, and ensuring that both genders are equally represented in teaching materials (J. F. K. Lee and Collins 2006) to avoid the formation of gender stereotypes.
2. **Enhanced career consultations to break gender stereotypes:** It could be beneficial for teachers to integrate career consultation into the curriculum more frequently, focusing on dismantling traditional gender stereotypes. By exposing students to diverse career paths and role models, educators can help students see their immense potential in different fields instead of confining themselves to arts/science subjects. Improved career consultations could then improve students' levels of career

exploration and amount of occupational information (Cheung and Jin 2016), helping them make informed choices in career decision making.

3. **Ongoing professional development for teachers:** This study recommends that teachers be provided with continuous professional development opportunities to stay updated on the latest research and effective strategies for motivating students in L2 learning. Training sessions can focus on gender-sensitive teaching methods, motivational techniques, and the integration of career planning into language instruction.

5.3. Limitations

Like any other systematic review, this meta-analysis comes with limitations. The most obvious limitation is the small sample size, as only eight eligible studies were included. This is because most motivational studies conducted in Hong Kong reported the results combined for both genders, but did not report the male and female statistics respectively. Due to limited data, a metaregression could not be carried out in the analysis of moderators, and a narrative synthesis of moderators was conducted instead. While a narrative synthesis could go hand-in-hand with statistical data to organize, describe, explore, and interpret the moderators of study findings (Thomson and Campbell 2020), it cannot estimate bivariate correlations while controlling for all other factors (J. P. T. Higgins et al. 2023). Furthermore, with the small sample size, there was missing data for moderators, especially school banding and L2 proficiency. With the limited data points, it was difficult to draw a statistically sound conclusion on the moderators' effect on gender differences in L2 motivation. This limitation points to a need for more studies on moderators of L2 motivation; as motivation is a dynamic construct that could be influenced by environmental factors (Ramezanzadeh 2021), investigations into the confounding factors of L2 motivation under specific cultural contexts would be most valuable.

Secondly, the current review only included studies written in English or Chinese (traditional and simplified), which might have limited the search results. However, since English and Chinese are the two major written languages used in Hong Kong, the author is confident that this review has captured the majority of research published in this area.

Lastly, due to the nature of meta-analyses, this study only included quantitative and mixed-methods studies, and left qualitative studies unexplored. As a result, it was difficult to qualitatively assess the experiential and social factors that gave rise to the observed gender

difference in L2 motivation. Future research could consider building on quantitative findings by exploring complex relationships and nuances through qualitative data.

5.4. Areas for Future Research

Summarizing the limitations identified, it is apparent that more longitudinal studies are required to capture the dynamic, fluctuational nature of L2 motivation. In the future, systematic reviews on motivation could also consider synthesizing findings from qualitative studies to add depth and explanatory power to statistical findings.

Moreover, more studies are required to investigate the confounding factors of L2 motivation. As motivation is highly dynamic and depends on the specific contexts the learners are in, it would be worthwhile to examine the interplay between learners' L2 motivation and environmental factors under different cultural contexts. This meta-analysis has only highlighted the effect of gender and four potential moderators, leaving much room for further explorations.

Lastly, methodologically, researchers should also pay heed to the compatibility between the motivational theories and instruments they employ to avoid the conflation, oversimplification, and mismatch of theories that have already pervaded this field. For studies using self-designed tools, Confirmatory Factor Analysis should also be conducted to ensure the validity and reliability of their instruments.

6. CONCLUSION

This meta-analysis examined the gender differences in L2 motivation among secondary school students in Hong Kong. A total of eight research reports, involving 24,660 language learners, were meta-analyzed. The eligible studies involved multiple frameworks for measuring motivation. Nonetheless, the data were combined as the criterion measures across studies were largely conflated. The combined results showed a significant but small gender difference in favour of females, indicating that they were slightly more motivated than their male counterparts in English language classrooms. The results also revealed a significant age effect, where the gender difference in L2 motivation significantly increased with age. The gender difference first became significant in Secondary 3, but became insignificant again in Secondary 6. No significant effect was observed for the medium of instruction, school banding, and L2 proficiency, possibly due to missing data. However, most of the studies involved methodological issues, hampering their validity and empirical strength to support claims on gender differences in L2 motivation.

This study provided crucial evidence for the criticality of adolescence in L2 motivation. Given the observed differentiation in Secondary 3, it is recommended that teachers devise focused motivational strategies, enhance career consultations, and participate in ongoing professional development to encourage L2 learners and dismantle gender stereotypes in the language classroom.

Future systematic reviews could consider synthesizing findings from qualitative studies to add explanatory power to any observed gender differences. More studies are also required to study the moderators of gender differences in L2 motivation. Most importantly, motivation researchers should be cautious with the adequacy of the motivational frameworks and instruments they employ to produce reliable evidence on the gender effect in L2 motivation.

Funding: This research received no external funding.

References

Alastair, Henry, and Meng Liu. 2023. “Can L2 Motivation Be Modelled as a Self-System? A Critical Assessment.” *System* 119 (December):1–13.

Al-Hoorie, Ali H. 2018. “The L2 Motivational Self System: A Meta-Analysis.” *Studies in Second Language Learning and Teaching* 8 (4): 721–54. https://doi.org/10.14746/ssllt.2018.8.4.2.

Au, Mei Yan Florence. 2008. “Investigating Gender in Students’ English Learning Beliefs in an English as a Second Language (ESL) Class.” MA thesis, Hong Kong: The University of Hong Kong. https://doi.org/10.5353/th_b4126205.

Azzouz Boudadi, Nadia, and Mar Gutiérrez-Colón. 2020. “Effect of Gamification on Students’ Motivation and Learning Achievement in Second Language Acquisition within Higher Education: A Literature Review 2011-2019.” *The EuroCALL Review* 28 (1): 40. https://doi.org/10.4995/eurocall.2020.12974.

Bacon, Susan M. C., and Michael D. Finnemann. 1992. “Sex Differences in Self-Reported Beliefs About Foreign-Language Learning and Authentic Oral and Written Input.” *Language Learning* 42 (4): 471–95. https://doi.org/10.1111/j.1467-1770.1992.tb01041.x.

Balshem, Howard, Mark Helfand, Holger J. Schünemann, Andrew D. Oxman, Regina Kunz, Jan Brozek, Gunn E. Vist, Yngve Falck-Ytter, Joerg Meerpohl, and Susan Norris. 2011. “GRADE Guidelines: 3. Rating the Quality of Evidence.” *Journal of Clinical Epidemiology* 64 (4): 401–6. https://doi.org/10.1016/j.jclinepi.2010.07.015.

Bećirović, Senad, and Remzija Hurić–Bećirović. 2017. “The Role of Age in Students’ Motivation and Achievement in Learning English as a Second Language.” *The Journal of Linguistic and Intercultural Education* 10 (1): 23–36. https://doi.org/10.29302/jolie.2017.10.1.2.

Belmechri, Faïza, and Kirsten Hummel. 1998. “Orientations and Motivation in the Acquisition of English as a Second Language Among High School Students in Quebec City.” *Language Learning* 48 (2): 219–44. https://doi.org/10.1111/1467-9922.00040.

Blair, Ryan, and Mahmoud Azaz. 2019. “Self-Guide Distinctions and Episodic Memory in L2 Motivation: The Case of Arabic in the United States.” *System* 81 (April):192–205. https://doi.org/10.1016/j.system.2019.01.015.

Boekaerts, Monique. 1995. *Motivation in Education*. Vernon-Wall Lecture 14. Education Section of the British Psychological Society.

Boo, Zann, Zoltán Dörnyei, and Stephen Ryan. 2015. “L2 Motivation Research 2005–2014: Understanding a Publication Surge and a Changing Landscape.” *System* 55 (December):145–57. https://doi.org/10.1016/j.system.2015.10.006.

Borenstein, Michael, Larry Hedges, and Hannah Rothstein. 2007. *Meta-Analysis Fixed Effect vs. Random Effects*. Meta-analysis.com.

Boyatzis, Richard E., and Kleio Akrivou. 2006. “The Ideal Self as the Driver of Intentional Change.” Edited by Richard Boyatzis. *Journal of Management Development* 25 (7): 624–42. https://doi.org/10.1108/02621710610678454.

Brutt-Griffler, Janina, and Sumi Kim. 2016. “Closing the Gender Gap: The Role of English.” In *English as a Lingua Franca: Perspectives and Prospects*, edited by Marie-Luise Pitzl and Ruth Osimk-Teasdale, 245–57. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9781501503177.

———. 2018. “Gender Socialization: From L1 to L2 Languacultures.” *International Journal of Applied Linguistics* 28 (1): 102–18. https://doi.org/10.1111/ijal.12186.

Census and Statistics Department, HKSAR. 2024. “Use of Language in Hong Kong 香港的語言使用情況.” *Hong Kong Monthly Digest of Statistics*, February, 1–15.

Chaffee, Kathryn Everhart, Nigel Mantou Lou, Kimberly A. Noels, and Joshua W. Katz. 2020. “Why Don’t ‘Real Men’ Learn Languages? Masculinity Threat and Gender Ideology Suppress Men’s Language Learning Motivation.” *Group Processes & Intergroup Relations* 23 (2): 301–18. https://doi.org/10.1177/1368430219835025.

Chalmers, Hamish, Jess Brown, and Anastasia Koryakina. 2023. “Topics, Publication Patterns, and Reporting Quality in Systematic Reviews in Language Education. Lessons from the International Database of Education Systematic Reviews (IDESR).” *Applied Linguistics Review*, January, 1–25. https://doi.org/10.1515/applirev-2022-0190.

Chan, Tat Wah Edmond. 2000. “A Case Study on the Motivational Pattern of Learning ESL in S.W.C.S. Chan Pak Sha School: In Application to Self Access Learning.” MA thesis, Hong Kong: Hong Kong Baptist University. https://scholars.hkbu.edu.hk/en/studentTheses/a-case-study-on-the-motivational-pattern-of-learning-esl-in-swcs-.

Chavez, Monika. 2000. “Teacher and Student Gender and Peer Group Gender Composition in German Foreign Language Classroom Discourse: An Exploratory Study.” *Journal of Pragmatics* 32 (7): 1019–58. https://doi.org/10.1016/S0378-2166(99)00065-X.

Cheung, Raysen, and Qiuping Jin. 2016. “Impact of a Career Exploration Course on Career Decision Making, Adaptability, and Relational Support in Hong Kong.” *Journal of Career Assessment* 24 (3): 481–96. https://doi.org/10.1177/1069072715599390.

Chung, Yin Shan. 2016. “Examining the Educational Bands System in Hong Kong and How It Impacts Students’ Self-Concept.” BSW thesis, Florida, United States: The Florida State University.

Clark, Ann, and John Trafford. 1995. “Boys into Modern Languages: An Investigation of the Discrepancy in Attitudes and Performance between Boys and Girls in Modern Languages.” *Gender and Education* 7 (3): 315–26. https://doi.org/10.1080/09540259550039022.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, N.J.: Erlbaum Associates. https://doi.org/10.4324/9780203771587.

Cortés, Kristin Hull. 2002. “Youth and the Study of Foreign Language: An Investigation of Attitudes.” *Foreign Language Annals* 35 (3): 320–32. https://doi.org/10.1111/j.1944-9720.2002.tb01857.x.

Csizér, Kata, and Judit Kormos. 2009. “Learning Experiences, Selves and Motivated Learning Behaviour: A Comparative Analysis of Structural Models for Hungarian Secondary and University Learners of English.” In *Motivation, Language Identity and the L2 Self*, edited by Zoltán Dörnyei and Ema Ushioda, 98–119. Multilingual Matters. https://doi.org/10.21832/9781847691293-006.

Csizér, Kata, and Gabriella Lukács. 2010. “The Comparative Analysis of Motivation, Attitudes and Selves: The Case of English and German in Hungary.” *System* 38 (1): 1–13. https://doi.org/10.1016/j.system.2009.12.001.

Deci, Edward L., and Richard M. Ryan. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4899-2271-7.

———, eds. 2004. *Handbook of Self-Determination Research*. Rochester, New York: University of Rochester Press.

Dörnyei, Zoltán. 1998. “Motivation in Second and Foreign Language Learning.” *Language Teaching* 31 (3): 117–35. https://doi.org/10.1017/S026144480001315X.

———. 2005. “Motivation and Self-Motivation.” In *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*, edited by Zoltán Dörnyei, 65–119. Routledge. https://doi.org/10.4324/9781410613349.

———. 2009. “The L2 Motivational Self System.” In *Motivation, Language Identity and the L2 Self*, edited by Zoltán Dörnyei and Ema Ushioda, 9–42. Multilingual Matters. https://doi.org/10.21832/9781847691293-005.

———. 2019. “Towards a Better Understanding of the L2 Learning Experience, the Cinderella of the L2 Motivational Self System.” *Studies in Second Language Learning and Teaching* 9 (1): 19–30. https://doi.org/10.14746/ssllt.2019.9.1.2.

Dörnyei, Zoltán, and Letty Chan. 2013. “Motivation and Vision: An Analysis of Future L2 Self Images, Sensory Styles, and Imagery Capacity Across Two Target Languages.” *Language Learning* 63 (3): 437–62. https://doi.org/10.1111/lang.12005.

Dörnyei, Zoltán, and Richard Clement. 2002. “Motivational Characteristics of Learning Different Target Languages Results of a Nationwide Survey.” In *Motivation and Second Language Acquisition*, edited by Zoltán Dörnyei, 2nd ed., 399–432. Technical Report / Second Language Teaching and Curriculum Center 23. Honolulu: University of Hawai’i Press.

Dörnyei, Zoltán, Kata Csizér, and Nóra Németh. 2006. “Modifying Factors in Language Attitudes and Motivation: Gender, Geographical Location and School Instruction.” In *Motivation, Language Attitudes and Globalisation*, 55–72. Multilingual Matters. https://doi.org/10.21832/9781853598876-005.

Dörnyei, Zoltán, and Ema Ushioda. 2013. “Motivation to Learn a Foreign/Second Language: A Historical Overview.” In *Teaching and Researching: Motivation*, 0 ed., 53–87. Routledge. https://doi.org/10.4324/9781315833750-11.

———. 2021. *Teaching and Researching Motivation*. Third edition. New York: Routledge.

Eccles, Jacquelynne. 1987. “Gender Roles and Achievement Patterns: An Expectancy Value Perspective.” In *Masculinity/Femininity: Basic Perspectives*, edited by June Machover Reinisch, Leonard A. Rosenblum, and June M. Reinisch, 240–80. The Kinsey Institute Series 1. New York: Oxford University Press.

Eccles, Jacquelynne, Allan Wigfield, Rena D. Harold, and Phyllis Blumenfeld. 1993. “Age and Gender Differences in Children’s Self- and Task Perceptions during Elementary School.” *Child Development* 64 (3): 830–47.

Fung, Kam Yin. 2006. “Sex Differences in English Learning in Junior Secondary School in Hong Kong.” EdD Thesis, Hong Kong: The Chinese University of Hong Kong. https://repository.lib.cuhk.edu.hk/en/item/cuhk-343847.

Gallagher, Matthew W., and Timothy A. Brown. 2013. “Introduction to Confirmatory Factor Analysis and Structural Equation Modeling.” In *Handbook of Quantitative Methods for Educational Research*, edited by Timothy Teo, 289–314. Rotterdam: SensePublishers. https://doi.org/10.1007/978-94-6209-404-8_14.

Gardner, Robert C. 1985a. *Social Psychology and Second Language Learning: The Role of Attitudes and Motivation*. The Social Psychology of Language 4. London: Edward Arnold.

———. 1985b. “The Attitude Motivation Test Battery Manual.” Manuscript. https://publish.uwo.ca/~gardner/docs/AMTBmanual.pdf.

———. 2001. “Integrative Motivation and Second Language Acquisition.” In *Motivation and Second Language Acquisition*, edited by Zoltán Dörnyei and Richard Schmidt, 1–20. Hawai’i: University of Hawai’i Press.

———. 2005. “Integrative Motivation and Second Language Acquisition.” In *Integrative Motivation and Second Language Acquisition*, 1–22. London, Canada. https://publish.uwo.ca/~gardner/docs/caaltalk5final.pdf.

———. 2006. “The Socio-Educational Model of Second Language Acquisition: A Research Paradigm.” *EUROSLA Yearbook* 6 (1): 237–60. https://doi.org/10.1075/eurosla.6.14gar.

———. 2007. “Motivation and Second Language Acquisition.” *Porta Linguarum* 8:9–20.

———. 2010a. *Motivation and Second Language Acquisition: The Socio-Educational Model*. Language as Social Action 10. New York: Peter Lang.

———. 2010b. “Second Language Acquisition: A Social Psychological Perspective.” In *The Oxford Handbook of Applied Linguistics*, edited by Robert B. Kaplan, 204–16. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195384253.013.0014.

Gardner, Robert C., R. N. Lalonde, and R. Pierson. 1983. “The Socio-Educational Model of Second Language Acquisition: An Investigation Using Lisrel Causal Modeling.” *Journal of Language and Social Psychology* 2 (1): 1–15. https://doi.org/10.1177/0261927X8300200101.

Gardner, Robert C., and Wallace E. Lambert. 1972. *Attitudes and Motivation in Second-Language Learning*. Rowley, Mass: Newbury.

Gardner, Robert C., and P. C. Smyihe. 1981. “On the Development of the Attitude/Motivation Test Battery.” *The Canadian Modern Language Review* 37 (3): 510–25. https://doi.org/10.3138/cmlr.37.3.510.

Gardner, Robert C., and Padric C. Smythe. 1975. “Second Language Acquisition: A Social Psychological Approach. Research Bulletin No. 332.”

Ghenghesh, Pauline. 2010. “The Motivation of L2 Learners: Does It Decrease with Age?” *English Language Teaching* 3 (1): p128. https://doi.org/10.5539/elt.v3n1p128.

Guyatt, Gordon H., Andrew D. Oxman, Regina Kunz, Jan Brozek, Pablo Alonso-Coello, David Rind, P. J. Devereaux, et al. 2011. “GRADE Guidelines 6. Rating the Quality of Evidence—Imprecision.” *Journal of Clinical Epidemiology* 64 (12): 1283–93. https://doi.org/10.1016/j.jclinepi.2011.01.012.

Guyatt, Gordon H., Andrew D. Oxman, Regina Kunz, James Woodcock, Jan Brozek, Mark Helfand, Pablo Alonso-Coello, Paul Glasziou, et al. 2011. “GRADE Guidelines: 7. Rating the Quality of Evidence—Inconsistency.” *Journal of Clinical Epidemiology* 64 (12): 1294–1302. https://doi.org/10.1016/j.jclinepi.2011.03.017.

Guyatt, Gordon H., Andrew D. Oxman, Regina Kunz, James Woodcock, Jan Brozek, Mark Helfand, Pablo Alonso-Coello, Yngve Falck-Ytter, et al. 2011. “GRADE Guidelines: 8. Rating the Quality of Evidence—Indirectness.” *Journal of Clinical Epidemiology* 64 (12): 1303–10. https://doi.org/10.1016/j.jclinepi.2011.04.014.

Guyatt, Gordon H., Andrew D. Oxman, Victor Montori, Gunn Vist, Regina Kunz, Jan Brozek, Pablo Alonso-Coello, et al. 2011. “GRADE Guidelines: 5. Rating the Quality of Evidence—Publication Bias.” *Journal of Clinical Epidemiology* 64 (12): 1277–82. https://doi.org/10.1016/j.jclinepi.2011.01.011.

Guyatt, Gordon H., Andrew D. Oxman, Shahnaz Sultan, Jan Brozek, Paul Glasziou, Pablo Alonso-Coello, David Atkins, et al. 2013. “GRADE Guidelines: 11. Making an Overall Rating of Confidence in Effect Estimates for a Single Outcome and for All Outcomes.” *Journal of Clinical Epidemiology* 66 (2): 151–57. https://doi.org/10.1016/j.jclinepi.2012.01.006.

Guyatt, Gordon H., Andrew D. Oxman, Gunn E. Vist, Regina Kunz, Yngve Falck-Ytter, Pablo Alonso-Coello, and Holger J. Schünemann. 2008. “GRADE: An Emerging Consensus on Rating Quality of Evidence and Strength of Recommendations.” *British Medical Journal* 336 (7650): 924–26. https://doi.org/10.1136/bmj.39489.470347.AD.

Ha, Marie-Paule. 2009. “Double Trouble: Doing Gender in Hong Kong.” *Signs: Journal of Women in Culture and Society* 34 (2): 423–49. https://doi.org/10.1086/591088.

Harrison, Reema, Benjamin Jones, Peter Gardner, and Rebecca Lawton. 2021. "Quality Assessment with Diverse Studies (QuADS): An Appraisal Tool for Methodological and Reporting Quality in Systematic Reviews of Mixed- or Multi-Method Studies." *BMC Health Services Research* 21 (1): 144. https://doi.org/10.1186/s12913-021-06122-y.

Heras, Arantxa, and David Lasagabaster. 2015. "The Impact of CLIL on Affective Factors and Vocabulary Learning." *Language Teaching Research* 19 (1): 70–88. https://doi.org/10.1177/1362168814541736.

Hessel, Gianna. 2015. "From Vision to Action: Inquiring into the Conditions for the Motivational Capacity of Ideal Second Language Selves." *System* 52 (August):103–14. https://doi.org/10.1016/j.system.2015.05.008.

Higgins, E. Tory. 1987. "Self-Discrepancy: A Theory Relating Self and Affect." *Psychological Review* 94 (3): 319–40. https://doi.org/10.1037/0033-295X.94.3.319.

Higgins, Julian P. T., James Thomas, Jacqueline Chandler, Miranda Cumpston, Tianjing Li, Matthew J. Page, and Vivian A. Welch, eds. 2023. *Cochrane Handbook for Systematic Reviews of Interventions Version 6.4*. Cochrane. www.training.cochrane.org/handbook.

Ho, Maxwell Chun Sing, and Daphnee Hui Lin Lee. 2020. "School Banding Effects on Student Financial Literacy Acquisition in a Standardised Financial Literacy Curriculum." *The Asia-Pacific Education Researcher* 29 (4): 377–91. https://doi.org/10.1007/s40299-019-00491-z.

Horwitz, Elaine K. 1988. "The Beliefs about Language Learning of Beginning University Foreign Language Students." *The Modern Language Journal* 72 (3): 283–94. https://doi.org/10.1111/j.1540-4781.1988.tb04190.x.

Huang, Xuechun, and Hamish Chalmers. 2023. "Implementation and Effects of Pedagogical Translanguaging in EFL Classrooms: A Systematic Review." *Languages* 8 (3): 194. https://doi.org/10.3390/languages8030194.

Huang, Ziwei. 2020. "Thirteen Years since the First iPhone: A Systematic Review on the Effectiveness of Language Learning Apps on Smart Devices." MSc dissertation, University of Oxford. https://solo.bodleian.ox.ac.uk/permalink/44OXF_INST/35n82s/alma991025233671007026.

Islam, Muhammad, Martin Lamb, and Gary Chambers. 2013. "The L2 Motivational Self System and National Interest: A Pakistani Perspective." *System* 41 (2): 231–44. https://doi.org/10.1016/j.system.2013.01.025.

Kalliyath, Shalini. 2015. "Gender as a Social Construct." In *Qualitative Research on Illness, Wellbeing and Self Growth: Contemporary Indian Perspectives*, edited by Kumar Ravi Priya and Ajit K. Dalal, 362. New Delhi: Routledge.

Kim, Tae-Young. 2009. "The Sociocultural Interface between Ideal Self and Ought-to Self: A Case Study of Two Korean Students' ESL Motivation." In *Motivation, Language*

Identity and the L2 Self, edited by Zoltán Dörnyei and Ema Ushioda, 274–94. Multilingual Matters. https://doi.org/10.21832/9781847691293-015.

Kim, Tae-Young, and Yoon-Kyoung Kim. 2014. "A Structural Model for Perceptual Learning Styles, the Ideal L2 Self, Motivated Behavior, and English Proficiency." *System* 46 (October):14–27. https://doi.org/10.1016/j.system.2014.07.007.

Kong, Ji Hyun, Jeong Eun Han, Sungjo Kim, Hunil Park, Yong Suk Kim, and Hyunjoo Park. 2018. "L2 Motivational Self System, International Posture and Competitiveness of Korean CTL and LCTL College Learners: A Structural Equation Modeling Approach." *System* 72 (February):178–89. https://doi.org/10.1016/j.system.2017.11.005.

Konno, Katsuyuki. 2011. "The Relationship Between L2 Selves, Intrinsic/Extrinsic Motivation and Motivated Behavior of Japanese EFL Learners." *Annual Review of English Language Education in Japan* 22 (March):345–60. https://doi.org/10.20581/arele.22.0_345.

Kormos, Judit, and Kata Csizér. 2008. "Age-Related Differences in the Motivation of Learning English as a Foreign Language: Attitudes, Selves, and Motivated Learning Behavior." *Language Learning* 58 (2): 327–55. https://doi.org/10.1111/j.1467-9922.2008.00443.x.

Kormos, Judit, Thom Kiddle, and Kata Csizér. 2011. "Systems of Goals, Attitudes, and Self-Related Beliefs in Second-Language-Learning Motivation." *Applied Linguistics* 32 (5): 495–516. https://doi.org/10.1093/applin/amr019.

Kruidenier, Bastian G., and Richard Clément. 1986. *The Effect of Context on the Composition and Role of Orientations in Second Language Acquisition*. Publication, B-157. Québec: Centre international de recherche sur le bilinguisme, International Center for Research on Bilingualism.

Kutuk, Gulsah, David W. Putwain, Linda K. Kaye, and Bethan Garrett. 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): 212–35. https://doi.org/10.1111/bjep.12446.

Lai, Mee-Ling. 2007. "Gender and Language Attitudes: A Case of Postcolonial Hong Kong." *International Journal of Multilingualism* 4 (2): 83–116. https://doi.org/10.2167/ijm068.0.

Lamb, Martin. 2004. "'It Depends on the Students Themselves': Independent Language Learning at an Indonesian State School." *Language, Culture and Curriculum* 17 (3): 229–45. https://doi.org/10.1080/07908310408666695.

———. 2012. "A Self System Perspective on Young Adolescents' Motivation to Learn English in Urban and Rural Settings." *Language Learning* 62 (4): 997–1023. https://doi.org/10.1111/j.1467-9922.2012.00719.x.

Lasagabaster, David. 2016. “The Relationship between Motivation, Gender, L1 and Possible Selves in English-Medium Instruction.” *International Journal of Multilingualism* 13 (3): 315–32. https://doi.org/10.1080/14790718.2015.1105806.

Lee, Icy, Shulin Yu, and Yuan Liu. 2018. “Hong Kong Secondary Students’ Motivation in EFL Writing: A Survey Study.” *TESOL Quarterly* 52 (1): 176–87. https://doi.org/10.1002/tesq.364.

Lee, Jackie F. K., and Peter Collins. 2006. “Gender Representation in Hong Kong English Textbooks.” In *Gender Representation in Hong Kong English Textbooks*, 1–40. Hong Kong: Hong Kong Institute of Education.

Lee, Kwok Wai. 2018. “Reading Motivation in Chinese and English in Hong Kong Secondary Schools Students: Influence of Schools and Demographic Factors.” EdD Thesis, Hong Kong: The Education University of Hong Kong. https://repository.eduhk.hk/en/publications/reading-motivation-in-chinese-and-english-in-hong-kong-secondary-.

Lee, Yeonsil, and Kyungja Ahn. 2013. “English Learning Motivation and English Academic Achievement of Korean Elementary School Students: The Effects of L2 Selves, International Posture, and Family Encouragement.” *Modern English Education* 14 (1): 127–52.

Lo, Yuen Yi, and Eric Siu Chung Lo. 2014. “A Meta-Analysis of the Effectiveness of English-Medium Education in Hong Kong.” *Review of Educational Research* 84 (1): 47–73. https://doi.org/10.3102/0034654313499615.

Maccoby, Eleanor Emmons, and Carol Nagy Jacklin. 1974. “Intellectual Abilities and Cognitive Styles.” In *The Psychology of Sex Differences*, 63–133. Stanford University Press.

MacIntyre, Peter D., Susan C. Baker, Richard Clément, and Leslie A. Donovan. 2003. “Sex and Age Effects on Willingness to Communicate, Anxiety, Perceived Competence, and L2 Motivation Among Junior High School French Immersion Students.” *Language Learning* 53 (S1): 137–66. https://doi.org/10.1111/1467-9922.00226.

Markus, Hazel, and Paula Nurius. 1986. “Possible Selves.” *American Psychologist* 41 (September):954–69.

Masgoret, Anne-Marie, and Robert C. Gardner. 2003. “Attitudes, Motivation, and Second Language Learning: A Meta–Analysis of Studies Conducted by Gardner and Associates.” *Language Learning* 53 (1): 123–63. https://doi.org/10.1111/1467-9922.00212.

McEown, Maya Sugita, Kimberly A. Noels, and Kathryn Everhart Chaffee. 2014. “At the Interface of the Socio-Educational Model, Self-Determination Theory and the L2 Motivational Self System Models.” In *TheImpact of Self-Concept on Language Learning*, edited by Kata Csizér and Michael Magid, 19–50. Multilingual Matters. https://doi.org/10.21832/9781783092383-004.

McEown, Maya Sugita, Yasuyo Sawaki, and Tetsuo Harada. 2017. “Foreign Language Learning Motivation in the Japanese Context: Social and Political Influences on Self.” *The Modern Language Journal* 101 (3): 533–47. https://doi.org/10.1111/modl.12411.

Meeus, Wim. 2011. “The Study of Adolescent Identity Formation 2000–2010: A Review of Longitudinal Research.” *Journal of Research on Adolescence* 21 (1): 75–94. https://doi.org/10.1111/j.1532-7795.2010.00716.x.

Moore, David S., George P. McCabe, and Bruce A. Craig. 2014. *Introduction to the Practice of Statistics*. 8th ed. New York: W.H. Freeman and Company, a Macmillan Higher Education Company.

Moskovsky, Christo, Turki Assulaimani, Silvia Racheva, and Jean Harkins. 2016. “The L2 Motivational Self System and L2 Achievement: A Study of Saudi EFL Learners.” *The Modern Language Journal* 100 (3): 641–54. https://doi.org/10.1111/modl.12340.

Murphy, P. Karen, and Patricia A. Alexander. 2000. “A Motivated Exploration of Motivation Terminology.” *Contemporary Educational Psychology* 25 (1): 3–53. https://doi.org/10.1006/ceps.1999.1019.

Niemiec, Christopher P., and Richard M. Ryan. 2009. “Autonomy, Competence, and Relatedness in the Classroom: Applying Self-Determination Theory to Educational Practice.” *Theory and Research in Education* 7 (2): 133–44. https://doi.org/10.1177/1477878509104318.

Noels, Kimberly A. 2002. “New Orientations in Language Learning Motivation: Towards a Model of Intrinsic, Extrinsic, and Integrative Orientations and Motivation.” In *Motivation and Second Language Acquisition*, edited by Zoltán Dörnyei and Richard W. Schmidt, 2nd ed. Technical Report / Second Language Teaching and Curriculum Center 23. Honolulu: University of Hawai’i Press.

Noels, Kimberly A., Luc G. Pelletier, Richard Clément, and Robert J. Vallerand. 2000. “Why Are You Learning a Second Language? Motivational Orientations and Self-Determination Theory.” *Language Learning* 50 (1): 57–85. https://doi.org/10.1111/0023-8333.00111.

Pae, Tae-Il. 2008. “Second Language Orientation and Self-Determination Theory: A Structural Analysis of the Factors Affecting Second Language Achievement.” *Journal of Language and Social Psychology* 27 (1): 5–27. https://doi.org/10.1177/0261927X07309509.

Page, Matthew J., David Moher, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, et al. 2021. “PRISMA 2020 Explanation and Elaboration: Updated Guidance and Exemplars for Reporting Systematic Reviews.” *British Medical Journal* 372 (160): 1–36. https://doi.org/10.1136/bmj.n160.

Pang, Tsz-yin. 2014. “An Investigation into Gender and Motivation Regarding English Language Acquisition in Secondary Schools in Hong Kong.” MA dissertation, Hong Kong: The University of Hong Kong. https://doi.org/10.5353/th_b5319503.

Papi, Mostafa. 2010. “The L2 Motivational Self System, L2 Anxiety, and Motivated Behavior: A Structural Equation Modeling Approach.” *System* 38 (3): 467–79. https://doi.org/10.1016/j.system.2010.06.011.

Papi, Mostafa, Anna Vitalyevna Bondarenko, Soheil Mansouri, Liying Feng, and Chen Jiang. 2019. “Rethinking L2 Motivation Research: The 2×2 Model of L2 Self-Guides.” *Studies in Second Language Acquisition* 41 (2): 337–61. https://doi.org/10.1017/S0272263118000153.

Papi, Mostafa, and Yasser Teimouri. 2012. “Dynamics of Selves and Motivation: A Cross-sectional Study in the EFL Context of Iran.” *International Journal of Applied Linguistics* 22 (3): 287–309. https://doi.org/10.1111/j.1473-4192.2012.00312.x.

Popay, Jennie, Helen Roberts, Amanda Sowden, Mark Petticrew, Lisa Arai, Mark Rodgers, and Nicky Britten. 2006. “Guidance on the Conduct of Narrative Synthesis in Systematic Reviews.” *The ESRC Methods Programme*, April, 1–92.

Powell, Robert C., and Julia D. Batters. 1985. “Pupils’ Perceptions of Foreign Language Learning at 12+: Some Gender Differences.” *Educational Studies* 11 (1): 11–23. https://doi.org/10.1080/0305569850110102.

Qi, Di. 2021. “Examining the L2 Motivational Self System through Meta-Analysis.” PhD thesis, Washington, D.C.: Georgetown University. https://repository.library.georgetown.edu/handle/10822/1064671.

Ramezanzadeh, Anna-Maria. 2021. “Motivation and Multiglossia: Exploring the Learning of Arabic in UK Schools.” DPhil thesis, Oxford: University of Oxford.

Ramezanzadeh, Anna-Maria, and Robert Woore. 2023. “A Scoping Review into the Teaching and Learning of Arabic as an Additional Language.” Qatar Foundation International. https://flare.shape404.agency/qfi-org/documents/c493ba68015742d1ade571997c139d90/Digital_Ramezanzadeh_and_Woore_with_QFI_-_Scoping_Review_Public_Report_09nov23.pdf.

Ravid, Ruth D. 1987. “The Correlation of Parents’ and Students’ Attitudes with Hebrew Language Achievement.” *Foreign Language Annals* 20 (5). https://www.proquest.com/docview/1311686075/citation/BBD14D57D86B4F8FPQ/1.

Rigby, C. Scott, Edward L. Deci, Brian C. Patrick, and Richard M. Ryan. 1992. “Beyond the Intrinsic-Extrinsic Dichotomy: Self-Determination in Motivation and Learning.” *Motivation and Emotion* 16 (3): 165–85. https://doi.org/10.1007/BF00991650.

Ryan, Richard M., and Edward L. Deci. 2000. “Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions.” *Contemporary Educational Psychology* 25 (1): 54–67. https://doi.org/10.1006/ceps.1999.1020.

———. 2013. “Toward a Social Psychology of Assimilation: Self-Determination Theory in Cognitive Development and Education.” In *Self-Regulation and Autonomy*, edited by Bryan W. Sokol, Frederick M. E. Grouzet, and Ulrich Müller, 1st ed., 191–207. Cambridge University Press. https://doi.org/10.1017/CBO9781139152198.014.

———. 2016. “Facilitating and Hindering Motivation, Learning, and Well-Being in Schools: Research and Observations from Self-Determination Theory.” In *Handbook of Motivation at School*, edited by Kathryn R. Wentzel and David B. Miele, 96–119. New York: Routledge.

———. 2017. *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. New York: Guilford Press.

Ryan, Stephen. 2009. “Self and Identity in L2 Motivation in Japan: The Ideal L2 Self and Japanese Learners of English.” In *Motivation, Language Identity and the L2 Self*, edited by Zoltán Dörnyei and Ema Ushioda, 120–43. Multilingual Matters. https://doi.org/10.21832/9781847691293-007.

Semmar, Yassir. 2006. “An Exploratory Study of Motivational Variables in a Foreign Language Learning Context.” PhD Thesis, Minnesota: Capella University.

Shaaban, Kassim A., and Ghazi Ghaith. 2000. “Student Motivation to Learn English as a Foreign Language.” *Foreign Language Annals* 33 (6). https://www.proquest.com/docview/1311686828/citation/7B17480975F4480EPQ/1.

Sung, Chit Cheung Matthew. 2022. “L2 Learning as Gendered Practices: Chinese Female Students’ English Learning Experiences at an English-Medium University in Hong Kong.” *System* 109 (October):102871. https://doi.org/10.1016/j.system.2022.102871.

Taguchi, Tatsuya, Michael Magid, and Mostafa Papi. 2009. “The L2 Motivational Self System among Japanese, Chinese and Iranian Learners of English: A Comparative Study.” In *Motivation, Language Identity and the L2 Self*, edited by Zoltán Dörnyei and Ema Ushioda, 66–97. Multilingual Matters. https://doi.org/10.21832/9781847691293-005.

Taie, Masumeh, and Asghar Afshari. 2015. “A Critical Review on the Socio-Educational Model of SLA.” *Theory and Practice in Language Studies* 5 (3): 605. https://doi.org/10.17507/tpls.0503.21.

Takahashi, Chika, and Seongah Im. 2020. “Comparing Self-Determination Theory and the L2 Motivational Self System and Their Relationships to L2 Proficiency.” *Studies in Second Language Learning and Teaching* 10 (4): 673–96. https://doi.org/10.14746/ssllt.2020.10.4.2.

Teimouri, Yasser. 2017. “L2 Selves, Emotions, and Motivated Behaviors.” *Studies in Second Language Acquisition* 39 (4): 681–709. https://doi.org/10.1017/S0272263116000243.

Thomson, Hilary, and Mhairi Campbell. 2020. “‘Narrative Synthesis’ of Quantitative Effect Data in Cochrane Reviews: Current Issues and Ways Forward.” Webinar presented at the “Narrative synthesis” of quantitative effect data in Cochrane reviews: current issues and ways forward, Glasgow. https://training.cochrane.org/sites/training.cochrane.org/files/public/uploads/Webinar%201%20Introduction%203Feb2020.pdf.

Thorsen, Cecilia, Alastair Henry, and Christina Cliffordson. 2020. “The Case of a Missing Person? The Current L2 Self and the L2 Motivational Self System.” *International*

Journal of Bilingual Education and Bilingualism 23 (5): 584–600. https://doi.org/10.1080/13670050.2017.1388356.

Watt, Helen M. G. 2004. "Development of Adolescents' Self-Perceptions, Values, and Task Perceptions According to Gender and Domain in 7th- through 11th-Grade Australian Students." *Child Development* 75 (5): 1556–74.

———. 2016. "Gender and Motivation." In *Handbook of Motivation at School*, edited by Kathryn R. Wentzel and David B. Miele. New York: Routledge.

Wen, Xiaohong. 1997. "Motivation and Language Learning with Students of Chinese [1]." *Foreign Language Annals* 30 (2): 235–51. https://doi.org/10.1111/j.1944-9720.1997.tb02345.x.

West, Candace, and Don H. Zimmerman. 2002. "Doing Gender." In *Doing Gender, Doing Difference: Inequality, Power, and Institutional Change*, edited by Sarah Fenstermaker and Candace West, 3–24. New York: Routledge.

Wheaton, Blair, Bengt Muthen, Duane F. Alwin, and Gene F. Summers. 1977. "Assessing Reliability and Stability in Panel Models." In *Sociological Methodology*, edited by David R. Heise, 8:84–136. San Francisco: Jossey-Bass. https://www.jstor.org/stable/270754?origin=crossref.

Wong, Ruth. 2010. "Carrot or Stick? An Investigation into Motivation Orientations in Learning English among Hong Kong Chinese Students." *Revista Brasileira de Linguística Aplicada* 10 (1): 71–87. https://doi.org/10.1590/S1984-63982010000100005.

Yang, Jin-Suk, and Tae-Young Kim. 2011. "The L2 Motivational Self-System and Perceptual Learning Styles of Chinese, Japanese, Korean, and Swedish Students." *English Teaching* 66 (1): 141–62. https://doi.org/10.15858/engtea.66.1.201103.141.

You, Chenjing (Julia), Zoltán Dörnyei, and Kata Csizér. 2016. "Motivation, Vision, and Gender: A Survey of Learners of English in China." *Language Learning* 66 (1): 94–123. https://doi.org/10.1111/lang.12140.

Yousefi, Moslem, and Mohammad Hadi Mahmoodi. 2022. "The L2 Motivational Self-system: A Meta-analysis Approach." *International Journal of Applied Linguistics* 32 (2): 274–94. https://doi.org/10.1111/ijal.12416.

Yung, Kevin Wai-Ho. 2019. "Exploring the L2 Selves of Senior Secondary Students in English Private Tutoring in Hong Kong." *System* 80 (February):120–33. https://doi.org/10.1016/j.system.2018.11.003.

Zhang, Bennan. 2011. "Gender Dissonance in Language Attitudes: A Case of Hong Kong." *International Journal of Arts and Sciences* 4 (18): 77–109.

Appendices

Appendix A. Data Extraction Sheet

Category	Item	Description
General Information	Date form completed (yyyy/mm/dd)	The date of completion of this form.
	Reference citation	The full reference of the study.
	Year of publication	The year when the study was published.
	Gender of first author	The gender of the article's first author, identified through an internet search of their pronouns and/or photographs.
	Database/information source	The database/information source through which this article was found.
	Publication type	Indicate the type of publication, e.g., journal article, conference paper, MA dissertation.
	Language	The language in which the publication was written.
	Source of funding	Identify the funding source and any potential conflicts of interest.
Setting	Context of study	Identify the schools/communities/districts in Hong Kong where this study was carried out. If the study was carried out in a school, identify the school's medium of instruction.
Theoretical Framework	Motivational theory applied (if applicable)	Identify the motivational theories applied in the outcome measure designs. For example, L2MSS could be applied to a questionnaire measuring "ought-to self," "ideal L2 self," and "L2 learning experiences."
	Motivational components examined	Identify the motivational constructs that were examined in the study. For example, a study operationalizing L2MSS might examine "ought-to self," "ideal L2 self," or "L2 learning experiences."
Methods	Research questions	Identify the research questions for the study.
	Study design	Method:

		Does the study involve quantitative or mixed methods?
		<u>Approach:</u> Is the study cross-sectional or longitudinal? For longitudinal studies, what was the duration of the study?
		<u>Study design:</u> For randomized trials, is it a parallel, factorial, crossover, or cluster design? For non-randomized trials, is the study synchronic or diachronic? For other types of empirical studies, what was the study design?
	Intervention (if applicable)	Identify intervention characteristics that may modify the effect of the intervention. What materials and procedures were used? Who provided the intervention (i.e., did the moderators have any qualifications)? When was the intervention delivered, and how frequently?
	Comparator (if applicable)	Was a control group present in this study? If so, how were they defined? If not, what comparisons were made between experimental groups?
	Measurement tool(s)	Identify the tools through which each outcome domain was measured (e.g., questionnaire, cloze test). Provide details for each tool.
		<u>Validation of questionnaire (if applicable):</u> Has the questionnaire been validated/piloted? If so, what are its reliability coefficients (e.g., Kappa, ICC, Pearson r, Cronbach's Alpha)?
		<u>Interview questions (if applicable):</u> What questions were being asked in the interviews?

Participants	Recruitment method	Identify the channels through which participants were recruited, e.g., snowballing, personal connections, advertisement, email invitations.
	Gender composition	Indicate the number of boys and girls in the study.
	Demographics/ Eligibility criteria	Identify the characteristics of participants at the beginning of the study (e.g., age, socio-economic status, ESL proficiency)
	Sample size	Identify the total number of participants, and the number of participants in each subgroup (if any).
	Attrition	Identify i) the number of participants at the beginning, ii) the number of participants at the end, and iii) the attrition rate.
Outcomes	ESL skill(s) assessed	Identify the ESL skill(s) assessed and the assessment tools used, e.g., verbal, spelling, grammatical, writing abilities.
	Aspect(s) of motivation assessed	Identify the aspect(s) of motivation assessed and the assessment tools used, e.g., intrinsic motivation, ideal self, ought-to self.
	Outcomes with scores	Summarize the data for each measure and each group, including mean scores, SD, variance, SEM, etc.
	Effect size	Identify the between-group differences/effect sizes. If the effect size is not indicated, I will do my own calculation.
	Limitations	Indicate the limitations of the study, as identified by the authors and myself.
	Conclusion	Identify the key conclusions put forward by the author(s).
Comments	Additional comments	Indicate any additional information that is crucial to reviewing the study.

Appendix B. Risk of Bias Assessment (QuADS) (copied from Harrison et al. 2021)

QuADS Critera	0	1	2	3
1. Theoretical or conceptual underpinning to the research	No mention at all.	General reference to broad theories or concepts that frame the study. e.g. key concepts were identified in the introduction section.	Identification of specific theories or concepts that frame the study and how these informed the work undertaken. e.g. key concepts were identified in the introduction section and applied to the study.	Explicit discussion of the theories or concepts that inform the study, with application of the theory or concept evident through the design, materials and outcomes explored. e.g. key concepts were identified in the introduction section and the application apparent in each element of the study design.
2. Statement of research aim/s	No mention at all.	Reference to what the sought to achieve embedded within the report but no explicit aims statement.	Aims statement made but may only appear in the abstract or be lacking detail.	Explicit and detailed statement of aim/s in the main body of report.
3. Clear description	No mention at all.	General description of	Description of research setting is	Specific description of the

of research setting and target population		research area but not of the specific research environment e.g. 'in primary care.'	made but is lacking detail e.g. 'in primary care practices in region [x]'.	research setting and target population of study e.g. 'nurses and doctors from GP practices in [x] part of [x] city in [x] country.'
4. The study design is appropriate to address the stated research aim/s	No research aim/s stated, or the design is entirely unsuitable e.g. a Y/N item survey for a study seeking to undertake exploratory work of lived experiences.	The study design can only address some aspects of the stated research aim/s e.g. use of focus groups to capture data regarding the frequency and experience of a disease.	The study design can address the stated research aim/s but there is a more suitable alternative that could have been used or used in addition e.g. addition of a qualitative or quantitative component could strengthen the design.	The study design selected appears to be the most suitable approach to attempt to answer the stated research aim/s.
5. Appropriate sampling to address the	No mention of the sampling approach.	Evidence of consideration of the sample required	Evidence of consideration of sample required to address the aim.	Detailed evidence of consideration of the sample required to

research aim/s		e.g. the sample characteristics are described and appear appropriate to address the research aim/s.	e.g. the sample characteristics are described with reference to the aim/s.	address the research aim/s. e.g. sample size calculation or discussion of an iterative sampling process with reference to the research aims or the case selected for study.
6. Rationale for choice of data collection tool/s	No mention of rationale for data collection tool used.	Very limited explanation for choice of data collection tool/s. e.g. based on availability of tool.	Basic explanation of rationale for choice of data collection tool/s. e.g. based on use in a prior similar study.	Detailed explanation of rationale for choice of data collection tool/s. e.g. relevance to the study aim/s, co- designed with the target population or assessments of tool quality.
7. The format and content of data collection tool is appropriate	No research aim/s stated and/or data collection tool not detailed.	Structure and/or content of tool/s suitable to address some aspects of the research aim/s or to address the aim/s superficially	Structure and/or content of tool/s allow for data to be gathered broadly addressing the stated aim/s but	Structure and content of tool/s allow for detailed data to be gathered around all relevant issues required to

to address the stated research aim/s		e.g. single item response that is very general or an open-response item to capture content which requires probing.	could benefit from refinement. e.g. the framing of survey or interview questions are too broad or focused to one element of the research aim/s.	address the stated research aim/s.
8. Description of data collection procedure	No mention of the data collection procedure.	Basic and brief outline of data collection procedure e.g. 'using a questionnaire distributed to staff'.	States each stage of data collection procedure but with limited detail or states some stages in detail but omits others e.g. the recruitment process is mentioned but lacks important details.	Detailed description of each stage of the data collection procedure, including when, where and how data was gathered such that the procedure could be replicated.
9. Recruitment data provided	No mention of recruitment data.	Minimal and basic recruitment data e.g. number of people invited who agreed to take part.	Some recruitment data but not a complete account e.g. number of people who were invited and agreed.	Complete data allowing for full picture of recruitment outcomes e.g. number of people approached, recruited, and who completed

				with attrition data explained where relevant.
10. Justification for analytic method selected	No mention of the rationale for the analytic method chosen.	Very limited justification for choice of analytic method selected. e.g. previous use by the research team.	Basic justification for choice of analytic method selected e.g. method used in prior similar research.	Detailed justification for choice of analytic method selected e.g. relevance to the study aim/s or comment around of the strengths of the method selected.
11. The method of analysis was appropriate to answer the research aim/s	No mention at all.	Method of analysis can only address the research aim/s basically or broadly.	Method of analysis can address the research aim/s but there is a more suitable alternative that could have been used or used in addition to offer a stronger analysis.	Method of analysis selected is the most suitable approach to attempt answer the research aim/s in detail e.g. for qualitative interpretative phenomenological analysis might be considered

				preferable for experiences vs. content analysis to elicit frequency of occurrence of events.
12. Evidence that the research stakeholders have been considered in research design or conduct	No mention at all.	Consideration of some the research stakeholders e.g. use of pilot study with target sample but no stakeholder involvement in planning stages of study design.	Evidence of stakeholder input informing the research. e.g. use of pilot study with feedback influencing the study design/conduct or reference to a project reference group established to guide the research.	Substantial consultation with stakeholders identifiable in planning of study design and in preliminary work e.g. consultation in the conceptualisation of the research, a project advisory group or evidence of stakeholder input informing the work.
13. Strengths and limitations critically discussed	No mention at all.	Very limited mention of strengths and limitations with omissions of many key issues.	Discussion of some of the key strengths and weaknesses of the study but not complete.	Thorough discussion of strengths and limitations of all aspects of study including design, methods, data collection tools,

		e.g. one or two strengths/limitations mentioned with limited detail.	e.g. several strengths/limitations explored but with notable omissions or lack of depth of explanation.	sample & analytic approach.

Appendix C. GRADE Assessment Tool

Criteria	Descriptor	Score	Score awarded
Study Design	A body of literature begins with an assumption of "high" certainty (J. P. T. Higgins et al. 2023).	4	4
Downgrading factors			
Risk of Bias	**No serious risk of bias** - Most information is from studies at low risk of bias.	0	
	Serious risk of bias - Most information is from studies at moderate risk of bias.	-1	
	Very serious risk of bias - Most information is from studies at high risk of bias.	-2	
Inconsistency	**No serious inconsistency** 1. Point estimates barely vary across studies. 2. Confidence intervals (CIs) show wide overlap in the forest plot. 3. The Q-test (statistical test for heterogeneity) shows a high *p*-value. 4. Small I^2 (0-40%; J. P. T. Higgins et al. 2023).	0	
	Serious inconsistency 1. Point estimates moderately vary across studies. 2. Confidence intervals (CIs) show moderate overlap in the forest plot. 3. The Q-test (statistical test for heterogeneity) shows a moderate *p*-value. 4. Moderate I^2 (30-60%; J. P. T. Higgins et al. 2023).	-1	

	Very serious inconsistency 1. Point estimates vary widely across studies. 2. Confidence intervals (CIs) show minimal or no overlap in the forest plot. 3. The Q-test (statistical test for heterogeneity) shows a low *p*-value. 4. Substantial/considerable I^2 (50-100%; J. P. T. Higgins et al. 2023).	-2	
Indirectness	**No serious indirectness** 1. Little to no differences exist between the populations studied across studies. 2. Little to no differences exist between the interventions across studies. 3. Little to no differences exist between the outcome measures across studies. 4. Head-to-head comparisons between two or more interventions are highly feasible.	0	
	Serious indirectness 1. Some differences exist between the populations studied across studies such that the magnitude of estimated effect might differ moderately. 2. Some differences exist between the interventions across studies. 3. Some differences exist between the outcome measures across studies. 4. Head-to-head comparisons between two or more interventions are somewhat feasible.	-1	
	Very serious indirectness 1. Substantial differences exist between the populations studied across studies such that the	-2	

	magnitude of estimated effect might differ substantially. 2. Substantial differences exist between the interventions across studies. 3. Substantial differences exist between the outcome measures across studies. 4. Head-to-head comparisons between two or more interventions are impossible.		
Imprecision	**No serious imprecision** 1. Large sample size (≥400 participants; (J. P. T. Higgins et al. 2023)). 2. Confidence Interval (CI) is satisfactorily narrow.	0	
	Serious imprecision 1. Modest sample size. 2. Confidence Interval (CI) is moderate.	-1	
	Very serious imprecision 1. Small sample size. 2. Confidence Interval (CI) is wide.	-2	
Publication bias	**No serious publication biases** - Egger's regression shows little to no publication biases.	0	
	Serious publication bias - Egger's regression shows moderate publication bias.	-1	
	Very serious publication bias - Egger's regression shows very serious publication bias.	-2	
Upgrading factors			
Large effect	**Absence of large effect** - Major concerns about other issues, e.g., risk of bias, precision, and publication	0	

	bias are present, and may have affected the estimated effect sizes.		
	Evidence of large effect - The effect sizes are substantially large in a fair number of studies. - Few to no major concerns about other issues, e.g., risk of bias, precision, and publication bias, that may have affected the estimated effect sizes.	+1	
	Evidence of very large effect - The effect sizes are substantially large in most studies, particularly if they occur over short periods of time. - Few to no major concerns about other issues, e.g., risk of bias, precision, and publication bias, that may have affected the estimated effect sizes.	+2	
Dose-response gradient	**No evidence of a gradient** - No evidence of a positive correlation between an intervention (if applicable) and motivational outcomes.	0	
	Evidence of a gradient - Evidence of a positive correlation between an intervention (if applicable) and motivational outcomes.	+1	
All plausible confounding	**No effect on effect** - It is unlikely that any unmeasured confounding variables would increase the estimated effects.	0	
	Would decrease the estimated effect	+1	

	- It is likely that unmeasured confounding variables have lowered the estimated effects.		
	Would suggest a spurious effect when no effect is observed - It is likely that unmeasured confounding variables are correlated with the outcome measures. - If no effect was observed despite the potential correlations between outcome measures and confounding variables, consider rating up the evidence.	+1	
		Total score	

Corresponding quality of cumulative evidence:

≤1	2	3	≥4
Very weak	weak	moderate	high

Appendix D. Data extracted from individual studies

Category	Item	Description
General Information	Date form completed (yyyy/mm/dd)	2024-05-15
	Reference citation	Au, Mei Yan Florence. 2008. "Investigating Gender in Students' English Learning Beliefs in an English as a Second Language (ESL) Class." MA thesis, Hong Kong: The University of Hong Kong. https://doi.org/10.5353/th_b4126205.
	Year of publication	2008
	Gender of first author	Female.
	Database/information source	HKU Scholars Hub.
	Publication type	MA thesis.
	Language	English.
	Source of funding	N/A
Setting	Context of study	41 Secondary 2 students from the same class in an EMI co-educational school. The school is a Band one school, with around 25 years of history. There is no elite class system in the school.
Theoretical Framework	Motivational theory applied (if applicable)	The questionnaire used was developed based on Horwitz's (1988) Beliefs about Language Learning Inventory (BALLI) survey. The motivational theory applied for the motivation items are unknown, even after examining the original article proposing BALLI. Based on my own judgement, I find that the three items on motivation align quite closely with Gardner and Lambert's (1972) socio-educational model, as it assesses learners' integrative (e.g., I want to speak English well... to make friends with [English-speaking people]) and instrumental

		orientations (e.g., If I use English well... I will have chances to receieve better education).
	Motivational components examined	The following is my own judgement: (1) Integrative orientation; (2) instrumental orientation.
Methods	Research questions	(1) Are there gender differences in learner beliefs towards English learning in an ESL classroom? Is there any particular relationship between gender and learner beliefs in the class? Do the findings support or reject previous studies of gender and learner beliefs? (2) Why do we have such results? What are the implications?
	Study design	Mixed-methods.
		Cross-sectional.
		Synchronic study using a 4-point Likert scale questionnaire. Five boys and girls were randomly selected for follow-up interviews, learner diaries, and lesson feedback.
	Intervention (if applicable)	N/A
	Comparator (if applicable)	N/A
	Measurement tool(s)	The 4-point Likert scale questionnaire used was developed based on Horwitz's (1988) Beliefs about Language Learning Inventory (BALLI) survey. Three items (items 21, 25, 29) measured motivation. The questionnaire was in English and the Chinese translations for some difficult words were provided.
		The modified questionnaire's reliability and validity was not tested.
		(1) What do you think about English language learning? What should we learn? How should we

		learn? Why should we learn? What do you think it's very important when learning English? Is it different from learning other subjects? If yes, how is it different from other subjects? What makes you think like this? (2) How do you learn English? What do you do to learn English? Do you use the strategies listed in the questionnaire? Why do you learn in this way? (3) What activities would you like to have in English lessons? How would you like to have those activities conducted? (4) In learning English, do you like to practise with others? Why? If yes, who are the others that you would like to practise with? Family? Classmates? Senior form students? Boys or girls? Teachers? Native English speakers? How would you like to practise with them? Do you like to receive comments from them? Why? How do you feel when you receive comments from them? Why? (5) Based on your observation, who do you think is very good at English in your class? Why do you think they are good at English? (6) How do you feel when you make mistakes in using English? Why do you feel like this? What do you do to deal with the feelings? How do you feel when you are asked to use English in class? (7) What do you think or feel about the statement "learners should have the ability to be in charge of their own learning e.g. finding own ways of practicing and learning skills, setting learning goals and planning their learning"? Is it right or wrong? Why?

		(8) What do you think or feel about the statement "girls are good at learning languages and arts and boys are good at learning maths and science."? Why?
Participants	Recruitment method	Not indicated.
	Gender composition	20 males and 21 females. The author wrongly wrote "21 boys and 21 girls" in the methodology chapter, although she did write that there were only "41 students" participating in the study. From the appendix, it was apparent that 20 boys and 21 girls answered the questionnaire.
	Demographics/ Eligibility criteria	Participants were Secondary 2 students (ages 13-14) from a Band One EMI co-educational school.
	Sample size	41 participants (20 males and 21 females)
	Attrition	Not indicated.
Outcomes	ESL skill(s) assessed	N/A
	Aspect(s) of motivation assessed	The following is my own judgement: (1) Integrative orientation; (2) instrumental orientation.
	Outcomes with scores	Motivation item scores [strongly disagree M, strongly disagree F, disagree M, disagree F, agree M, agree F, strongly agree M, strongly agree F]: Item 21: 0, 0, 2, 0, 6, 6, 12, 15 Item 25: 0, 0, 3, 0, 9, 8, 8, 13 Item 29: 0, 0, 4, 3, 13, 10, 3, 7 The following is my own calculation: (1) Mean motivation for males (n=20, M=3.233, SD=0.708); (2) Mean motivation for females (n=21, M=3.511, SD=0.602); (3) Overall SD of population=0.671
	Effect size	The following is my own calculation: (1) ES for population= -0.414

		(2) Standard error of ES estimate=0.104
	Limitations	Limitations identified by the author: (1) The sample size was small; (2) the age group was limited to Secondary 2 students; (3) More data needs to be collected from teachers, parents, and from schools of different gender settings and bandings. (4) Only three items were used to measure students' motivation. Limitations identified by me: (1) The questionnaire used in the study has not been piloted and validated. (2) Important statistical data, such as SD, SEM, and *p*-value, were not reported. (3) In item 29, only 20 reponses from female students were recorded, although it was stated that 21 female students participated in the study. The reason was not stated. (4) The theoretical base of the author's collection and analysis of students' L2 motivation was not stated.
	Conclusion	(1) This study found that males scored higher than females in learner beliefs, which contradicts existing literature.
Comments	Additional comments	N/A

Category	Item	Description
General Information	Date form completed (yyyy/mm/dd)	2024-05-17
	Reference citation	Chan, Tat Wah Edmond. 2000. "A Case Study on the Motivational Pattern of Learning ESL in S.W.C.S. Chan Pak Sha School: In Application to Self Access Learning." MA thesis, Hong Kong: Hong Kong Baptist University. https://scholars.hkbu.edu.hk/en/studentTheses/a-case-study-on-the-motivational-pattern-of-learning-esl-in-swcs-.
	Year of publication	2000
	Gender of first author	Male.
	Database/information source	HKBU Scholars (student theses).
	Publication type	MA thesis.
	Language	English.
	Source of funding	N/A
Setting	Context of study	Participants were junior secondary school students (F1 to F3) from S.W.C.S. Chan Pak Sha School (CPS). CPS is a government-aided secondary grammar school in the Southern District of Hong Kong. The students belong to band 2-3. In the academic year 1999-2000, the F3 students were taught in English, while the F1 and F2 students were taught in Cantonese.
Theoretical Framework	Motivational theory applied (if applicable)	Ryan and Deci's (2000) self-determination theory.
	Motivational components examined	(1) intrinsic motivation; (2) extrinsic motivation.
Methods	Research questions	(1) Are the junior students in CPS school motivated in learning ESL?

		(2) What is the motivational pattern of the junior students in learning ESL? (3) What are the ways that the junior students like to learn? (4) How SALL can be put forward in CPS school?
	Study design	Mixed-methods.
		Cross-sectional.
		Synchronic study using a 5-point Likert scale adopted from Kruidenier and Clements (1986) and Belmechri and Hummel (1998).
	Intervention (if applicable)	N/A
	Comparator (if applicable)	N/A
	Measurement tool(s)	A 5-point questionnaire was adopted from Kruidenier and Clements (1986) and Belmechri and Hummel (1998). 7 items (items 1, 2, 3, 4, 5, 9, 12) were used to measure intrinsic motivation, and 5 items (items 6, 7, 8, 10, 11) were used to measure extrinsic motivations.
		The questionnaire was not piloted/validated in the present study.
		The discussion topic of the interview was "Why do you learn English?". Reponses were collected in four categories: "communicative purpose," "cultural curiosity," "academic concern," and "career concern."
Participants	Recruitment method	Quota sampling was used.
	Gender composition	41 males and 34 females.
	Demographics/ Eligibility criteria	Participants spoke Cantonese as their L1, and were enrolled in the English Language course of CPS.
	Sample size	75 participants (41 males and 34 females).
	Attrition	0%.
Outcomes	ESL skill(s) assessed	N/A

	Aspect(s) of motivation assessed	(1) intrinsic motivation; (2) extrinsic motivation.
	Outcomes with scores	Motivation scores [F1 male mean, F1 male SD, F1 female mean, F1 female SD, F2 male mean, F2 male SD, F2 female mean, F2 female SD, F3 male mean, F3 male SD, F3 female mean, F3 female SD]: Statement 1: 2.79, 1.58, 3.00, 1.00, 3.00, 1.07, 3.12, 0.93, 1.74, 1.15, 2.50, 0.75 Statement 2: 3.71, 1.27, 3.91, 0.94, 4.13, 0.35, 4.06, 0.97, 2.26, 1.28, 2.67, 1.52 Statement 4: 2.79, 1.12, 2.73, 1.10, 2.38, 1.19, 2.65, 1.06, 1.79, 0.88, 1.67, 0.82 Statement 9: 2.83, 1.17, 2.73, 2.04, 2.38, 2.19, 3.12, 0.99, 2.11, 1.92, 1.83, 2.00 Statement 12: 3.36, 1.15, 2.73, 1.10, 2.75, 0.89, 2.88, 1.93, 1.79, 1.13, 2.17, 1.17 Statement 3: 4.14, 0.95, 4.64, 0.50, 4.00, 0.93, 4.18, 0.88, 2.47, 1.43, 2.83, 1.33 Statement 5: 3.57, 1.16, 3.64, 1.43, 3.75, 0.89, 3.71, 0.59, 2.47, 1.22, 2.67, 1.63 Statement 7: 3.57, 0.94, 4.27, 0.65, 3.50, 0.93, 3.76, 0.83, 3.53, 1.50, 3.67, 1.03 Statement 11: 3.93, 1.14, 4.00, 0.89, 3.63, 1.06, 4.00, 0.87, 3.16, 1.50, 2.33, 1.51 Statement 6: 3.79, 1.05, 4.27, 0.79, 4.00, 1.07, 3.71, 1.53, 3.26, 1.41, 3.17, 0.75 Statement 8: 3.50, 1.22, 3.45, 1.04, 3.38, 1.6, 3.82, 0.95, 2.84, 1.38, 2.50, 0.84 Statement 10: 3.21, 1.19, 3.82, 0.75, 3.75, 0.89, 3.82, 0.81, 3.26, 1.37, 2.83, 0.98 The following is my own calculation:

		(1) Average motivation for males (n=41, M=3.018, SD=1.410); (2) Average motivation for females (n=34, M=3.403, SD=1.258); (3) Overall SD of population=1.356.
	Effect size	(1) ES for population = -0.284. (2) Standard error of ES estimate = 0.157
	Limitations	Limitations identified by me: (1) In the questionnaire, there are 7 items for intrinsic motivation, and 5 items for extrinsic motivation. This might lead to an imbalanced account of one's overall motivation. (2) The questionnaire was not validated or piloted with local secondary school students. (3) Some of the items are highly similar. For example, item 6 ("I work hard in learning English because it helps me to get a good job.") and item 10 ("I work hard in learning English because it helps me to develop my future career.") essentially have the same meaning.
	Conclusion	(1) F1 students were the most motivated, and F3 students were the least motivated in L2 learning among the three forms. (2) All students had similar views toward learning English in career and academic aspects. F1 students had higher motivation to learn English to communication, F2 students had higher motivation to know more about the culture of the English-speaking group. (3) F1 students prefer SALL to teacher-centred classroom learning, while F2 and F3 students had opposite views.
Comments	Additional comments	N/A

Category	Item	Description
General Information	Date form completed (yyyy/mm/dd)	2024-05-15
	Reference citation	Fung, Kam Yin. 2006. “Sex Differences in English Learning in Junior Secondary School in Hong Kong.” EdD Thesis, Hong Kong: The Chinese University of Hong Kong. https://repository.lib.cuhk.edu.hk/en/item/cuhk-343847.
	Year of publication	2006
	Gender of first author	Female.
	Database/information source	CUHK LibrarySearch (CUHK Theses).
	Publication type	EdD Thesis.
	Language	English.
	Source of funding	Fung's (2006) thesis was not funded. However, the original data set her study drew on was commissioned by the Education Department in Hong Kong.
Setting	Context of study	20,986 students from 101 schools participated in the study. The schools were EMI, CHI, CMI, and CLO, but the proportion is unknown.
Theoretical Framework	Motivational theory applied (if applicable)	A simplified version of Gardner's socio-educational model, consisting of "learning orientation" (further divided into integrative and instrumental orientations), "learning motivation" (further divided into motivation intensity, desire to learn, attitude toward learning English, and attitude toward bilingualism) and "learning achievement."
	Motivational components examined	(1) Motivation intensity; (2) desire to learn; (3) attitude toward learning English;

		(4) attitude toward bilingualism.
Methods	Research questions	(1) Are there sex differences in the academic achievements of English language among secondary-three students in Hong Kong? (2) Are there sex differences in the learning process of English language between boys and girls in secondary-three students in Hong Kong? More specifically, do Hong Kong secondary 3 students differ by sex in their learning orientation, motivation for English learning, motivation intensity, desire to learn English, attitudes toward learning English, and attitude toward bilingualism? (3) If sex differences are found in the above two research questions, then I may further the investigation by asking whether there are structural differences in the models accounting for the sex difference in English language achievement among Hong Kong secondary 3 students? More specifically, are there significant differences in covariance matrix, factor structure, factor loadings, error variances/covariances, factor variance/covariances, factor means and causal relationships between the male model and the female model?
	Study design	Quantitative.
		Cross-sectional.
		Synchronic study using a dataset involving a 5-point Likert scale questionnaire.
	Intervention (if applicable)	N/A
	Comparator (if applicable)	N/A
	Measurement tool(s)	A 5-point Likert scale questionnaire with 47 items was used. The questionnaire modified Gardner's

		Atittude/Motivation Test Battery, and measured "learning orientation" (10 items), "attitude toward learning English" (10 items), "attitude toward bilingualism" (12 items), "motivation intensity" (7 items), and "desire to learn" (8 items). The questionnaire was translated into Chinese to aid S3 students' comprehension.
		A Confirmatory Factor Analysis was conducted to establish the predictive relationships between parameters.
		N/A
Participants	Recruitment method	The data set for this study is borrowed from the "Evaluation on the Implementation of the Medium of Instruction Guidance for Secondary Schools" (MOI project), a research conducted by the Chinese University of Hong Kong and commissioned by the Education Department. The project aimed to monitor the academic and personal developments of junior secondary school students for three years. All Secondary 1 students in the 1999/2000 cohort were invited to sit in academic tests and fill in questionnaires. The current study only reviews the data from the 1999/2000 cohort when they were in Secondary 3 (academic year 2001/2002).
	Gender composition	11,052 males and 9,934 females.
	Demographics/ Eligibility criteria	Participants were Hong Kong secondary school students studying Secondary 3 (ages 14-16).
	Sample size	20,986 participants (11,052 males and 9,934 females).
	Attrition	Not indicated.
Outcomes	ESL skill(s) assessed	A English language test assessing students' listening, language form and function (e.g. vocabulary, syntax and grammar), reading comprehension, and writing

		skills. The test has a Cronbach's alpha value of 0.78, and a discrimination index of 0.37.
	Aspect(s) of motivation assessed	(1) Instrumental orientation; (2) integrative orientation; (3) attitude toward learning English; (4) attitude toward bilingualism; (5) motivation intensity; (6) desire to learn.
	Outcomes with scores	(1) Motivation items scores [Female N, female mean, female SD, male N, male mean, male SD]: Item 15: 6688, 1.84, 0.97, 6686, 2.14, 1.16 Item 16: 7329, 3.11, 1.10, 7279, 2.91, 1.14 Item 17: 7328, 4.12, 0.93, 7283, 3.92, 1.07 Item 18: 7325, 3.47, 0.89, 7279, 3.32, 1.01 Item 19: 7327, 3.65, 0.96, 7285, 3.50, 1.10 Item 20: 7323, 3.87, 0.90, 7271, 3.67, 1.04 Item 21: 7328, 3.86, 0.89, 7286, 3.68, 1.03 Item 22: 7330, 3.22, 0.90, 7286, 3.13, 1.03 Item 23: 7329, 4.17, 0.85, 7283, 3.96, 1.01 Item 24: 7325, 2.96, 0.98, 7275, 2.94, 1.10 Item 25: 7326, 4.20, 0.93, 7282, 4.00, 1.02 Item 26 (Reverse): 7334, 2.54, 0.93, 7287, 2.73, 1.08 Item 27: 7335, 3.94, 0.89, 7285, 3.81, 1.00 Item 28: 7335, 4.12, 0.84, 7284, 3.96, 0.99 Item 29: 7323, 4.15, 0.83, 7281, 3.99, 0.98 Item 30 (Reverse): 7326, 2.46, 1.02, 7275, 2.72, 1.16 Item 31: 7334, 3.98, 0.89, 7292, 3.84, 1.04 Item 32: 7330, 3.40, 0.87, 7284, 3.24, 0.98 Item 33: 7321, 3.29, 0.95, 7283, 3.21, 1.06 Item 34: 7319, 3.29, 0.95, 7278, 3.17, 1.06 Item 35: 7321, 3.62, 1.00, 7265, 3.58, 1.09 Item 36 (Reverse): 7330, 2.39, 1.09, 7289, 2.64, 1.20 Item 37: 7329, 4.13, 0.83, 7283, 3.91, 0.98

		Item 38 (Reverse): 7319, 2.71, 0.89, 7276, 2.77, 1.03 Item 39 (Reverse): 7324, 2.50, 0.98, 7276, 2.71, 1.12 Item 40 (Reverse): 7322, 3.31, 1.19, 7270, 3.50, 1.21 Item 41: 7329, 3.96, 0.90, 7286, 3.78, 1.05 Item 42 (Reverse): 7328, 3.13, 1.09, 7283, 3.10, 1.14 Item 43: 7326, 4.06, 0.86, 7279, 3.90, 1.00 Item 44 (Reverse): 7322, 1.93, 0.93, 7269, 2.24, 1.09 Item 45: 7318, 3.72, 0.92, 7277, 3.56, 1.06 Item 46 (Reverse): 7325, 2.63, 1.08, 7283, 2.77, 1.19 Item 47: 7321, 2.58, 1.19, 7278, 2.84, 1.22 Item 48 (Reverse): 7321, 2.64, 0.99, 7283, 2.86, 1.10 Item 49: 7321, 3.75, 0.89, 7275, 3.59, 1.03 Item 50 (Reverse): 7327, 3.90, 0.94, 7283, 3.83, 1.06 Item 51: 7328, 3.36, 1.21, 7289, 3.10, 1.28 Item 52 (Reverse): 7331, 2.16, 0.94, 7279, 2.41, 1.04 Item 53: 7319, 3.70, 0.94, 7270, 3.57, 1.06 Item 54: 7317, 2.36, 0.98, 7268, 2.41, 1.07 Item 55 (Reverse): 7320, 2.63, 0.99, 7272, 2.84, 1.11 Item 56: 7333, 3.08, 1.11, 7283, 2.88, 1.16 Item 57: 7332, 4.09, 0.88, 7277, 3.91, 1.01 Item 58 (Reverse): 7321, 3.05, 0.99, 7281, 3.23, 1.08 Item 59: 7323, 2.34, 1.04, 7276, 2.64, 1.15 Item 60 (Reverse): 7322, 3.02, 1.05, 7251, 3.19, 1.15 Item 61: 7329, 3.43, 1.05, 7283, 3.19, 1.15 The following is my own calculation: (1) Overall motivation for males (n=11,052, M=2.996, SD=1.329) (2) Overall motivation for females (n=9,934, M=3.126, SD=1.264) (3) Overall SD of population=1.298 (4) Overall mean of population=3.061
	Effect size	The following is my own calculation:

		(1) ES for population= -0.100 (2) Standard error of ES estimate=0.009
	Limitations	Limitation identified by the author: (1) The study did not evaluate the longitudinal development of students' L2 motivation, even though the data set should have enabled the author to do so. Limitations identified by me: (1) The author made no justification for choosing S3 as the group of focus. (2) Since the motivation questionnaire used in this study was modified based on Gardner's AMTB, the author should provide evidence on its reliability (e.g., Cronbach's alpha). (3) The parameters of L2 motivation are weighted differently in the questionnaire. For example, there are only seven items measuring motivational intensity, but there are 12 items measuring students' attitude toward learning English. When calculating a mean motivation score, this may lead to a biased result.
	Conclusion	(1) Boys have poorer English achievement. (2) Boys also have significantly lower factor means in instrumental orientation, integrative orientation, motivation intensity (in formal settings and English lessons), attitudes towards learning English, and attitudes towards bilingualism. (3) The causal relation between instrumental orientation and motivation for English learning is higher in boys.
Comments	Additional comments	(1) Confirmatory factor analysis was conducted to establish the motivation model presented in the study.

		(2) The study was very large in scale, covering about 25% of all secondary schools in Hong Kong at the time. It implies that the findings from the study are highly generalizable. (3) The original raw data set was unobtainable. I have emailed the Hong Kong Institute of Educational Research to obtain the original data set, but since the original author has retired a long time ago, they were unable to retrieve the data.

Category	Item	Description
General Information	Date form completed (yyyy/mm/dd)	2024-05-12
	Reference citation	Lai, Mee-Ling. 2007. "Gender and Language Attitudes: A Case of Postcolonial Hong Kong." *International Journal of Multilingualism* 4 (2): 83–116. https://doi.org/10.2167/ijm068.0.
	Year of publication	2007
	Gender of first author	Female.
	Database/information source	ProQuest Social Science Premium Collection.
	Publication type	Journal article.
	Language	English.
	Source of funding	N/A
Setting	Context of study	564 student participants came from Chinese Medium Instruction (CMI) schools, and 448 came from EMI schools. The schools' bandings were not specified. 28 teachers were also recruited from different schools.
Theoretical Framework	Motivational theory applied (if applicable)	Socio-educational model.
	Motivational components examined	(1) Integrative orientation toward English; (2) Instrumental orientation toward English. Other irrelevant components: (1) Integrative orientation toward Cantonese; (2) Instrumental orientation toward Cantonese; (3) Integrative orientation toward Putonghua; (4) Instrumental orientation toward Putonghua.
Methods	Research questions	(1) Do male and female students differ significantly in their integrative orientation towards Cantonese, English and Putonghua?

		(2) Do male and female students differ significantly in their instrumental orientation towards Cantonese, English and Putonghua?
	Study design	Mixed-methods.
		Cross-sectional.
		Synchronic study using a 4-point Likert scale questionnaire. An interview was also conducted with 19 male students and 21 female students from different schools.
	Intervention (if applicable)	N/A
	Comparator (if applicable)	N/A
	Measurement tool(s)	A 4-point Likert scale questionnaire with 24 items was used. The questionnaire measured six parameters, as indicated in "motivational components measured."
		A Confirmatory Factor Analysis was conducted to validate the six parameters. It generated a Goodness-of-fit Index (GFI)=0.91 and Normed Fit Index (NFI)=0.9. Moreover, the Cronbach's alphas of all constructs fall between 0.6 and 0.83, indicating moderate to high internal reliability.
		In order to avoid standard answers from students, the researcher asked indirect questions to investigate students' attitudes towards Cantonese, English, and Putonghua. For example, students were asked whether they would like to be an English-speaker, Cantonese-speaker or Putonghua-speaker if they were given a choice. Other sample questions include "Which language do you like best?", "Do you think speaking fluent English is cool?", and "Do you think speaking fluent Putonghua is cool?"

Participants	Recruitment method	Convenience sampling was used to recruit teachers (i.e., the researcher recruited teachers attending a professional development seminar).
	Gender composition	555 male (365 CMI, 190 EMI) and 493 female (199 CMI, 294 EMI).
	Demographics/ Eligibility criteria	The participants were the first cohort of secondary school students under the Mother Tongue policy. All participants were Secondary 4 students, aged 14-16.
	Sample size	1,048 participants (555 males and 493 females).
	Attrition	Not indicated.
Outcomes	ESL skill(s) assessed	N/A
	Aspect(s) of motivation assessed	(1) Integrative orientation toward English (measured by 6 4-point Likert scale items); (2) Instrumental orientation toward English (measured by 4 4-point Likert scale items).
	Outcomes with scores	(1) English integrative motivation for males (n=555, M=2.97, SD=0.47); (2) English instrumental motivation for males (n=555, M=3.46, SD=0.40); (3) English integrative motivation for females (n=493, M=3.14, SD=0.40); (4) English instrumental motivation for females (n=493, M=3.56, SD=0.26) The following is my own calculation: (5) Overall English motivation for male students (n=555, M=3.215, SD=0.501) (6) Overall English motivation for female students (n=986, M=3.350, SD=0.397) (7) Overall SD of entire population=0.460 (8) Overall English motivation for all students=3.301
	Effect size	(1) ES for English integrative motivation = -0.40 (direction standardized);

		(2) ES for English instrumental motivation = -0.30 (direction standardized); The following is my own calculation: (3) ES for overall motivation= -0.293 (3) Standard error of ES estimate=0.011
	Limitations	Limitations identified by the author: (1) The gender difference might be an effect of MOI; (2) The sample size for interviews was small. Limitations identified by me: (1) The study falsely interpreted Gardner's socio-educational model (wrongly labelled as ""sociocultural model"" in this study) as an integrative-instrumental dichotomy.
	Conclusion	(1) Male participants had more positive attitudes toward vernacular languages, and female participants had more positive attitudes toward non-native languages. (2) The gender differences in language attitudes is small.
Comments	Additional comments	N/A

Category	Item	Description
General Information	Date form completed (yyyy/mm/dd)	2024-05-16
	Reference citation	Lee, Icy, Shulin Yu, and Yuan Liu. 2018. "Hong Kong Secondary Students' Motivation in EFL Writing: A Survey Study." *TESOL Quarterly* 52 (1): 176–87. https://doi.org/10.1002/tesq.364.
	Year of publication	2018
	Gender of first author	Female.
	Database/information source	ProQuest Social Science Premium Collection.
	Publication type	Journal article.
	Language	English.
	Source of funding	N/A
Setting	Context of study	1,395 students (696 males and 699 females) across three different grades (Secondary 1 = 386, Secondary 3 = 500, and Secondary 5 = 509) from three secondary schools (Band 1 = 471, Band 2 = 452, and Band 3 = 472) in Hong Kong.
Theoretical Framework	Motivational theory applied (if applicable)	(1) Dörnyei's (2005) L2 Motivational Self-system; (2) Dörnyei's () Three-Level Framework; (3) Troia et al's (2013) writing motivation survey.
	Motivational components examined	(1) Importance of writing; (2) ideal L2 self; (3) ought-to self; (4) interest in writing; (5) self-regulation; (6) writing efficacy; (7) classroom writing context.
Methods	Research questions	(1) The extent to which Hong Kong secondary students are motivated to write in English; (2) the influence of language proficiency, gender, and grade on their writing motivation.

	Study design	Quantitative.
		Cross-sectional.
		Synchronic study using a 5-point Likert scale questionnaire. The descriptive statistics of the data is standardized to a +-1 scale (confirmed with authors).
	Intervention (if applicable)	N/A
	Comparator (if applicable)	N/A
	Measurement tool(s)	A 5-point Likert scale questionnaire was created from existing L2 learning motivation and first language writing motivation scales (Dörnyei, 2005; Dornyei & Csizer, 2002; Dörnyei & Ushioda, 2011; Troia et al., 2013). 45 items were present in the first draft, and there were 40 items in the final version after CFA. The questionnaire measured: (1) importance of writing (language level) (items 1-8); (2) ideal L2 self (items 9-11); (3) ought-to L2 self (items 12-15); (4) interest in writing (learner level) (items 16-17); (5) Self-regulation (learner level) (items 18-20); (6) Writing efficacy (learner level) (items 21-30); and (7) Classroom writing context (learning situation level/L2 language experience) (items 31-40).
		The first draft of the questionnaire was reviewed by researchers and a group of doctoral students in the field of L2 education. The researchers also conducted CFA, and removed five items due to weak factor loadings. The CFA for the remaining 40 items produced strong fit indices (x^2 = 7616.632, df= 719, CFI= 0.907, RMSEA=0.083, TLI=0.899). The questionnaire also had good internal consistency (a=0.803-0.933), except the OTS subscale (a=0.657).
		N/A

Participants	Recruitment method	Not indicated.
	Gender composition	696 males and 699 females.
	Demographics/ Eligibility criteria	Participants were Secondary 1, Secondary 3, and Secondary 5 students from Band 1, 2, and 3 schools. In Hong Kong, Band 1 schools usually accommodate students with the strongest academic abilities, followed by 2, and 3 (weakest).
	Sample size	1,395 students (696 males and 699 females)
	Attrition	Not indicated.
Outcomes	ESL skill(s) assessed	Students' motivation for English writing was assessed. The aforementioned questionnaire was used.
	Aspect(s) of motivation assessed	(1) Importance of writing; (2) ideal L2 self; (3) ought-to self; (4) interest in writing; (5) self-regulation; (6) writing efficacy; (7) classroom writing context.
	Outcomes with scores	Motivation scores [importance of writing, ideal L2 self, ought-to L2 self, interest in writing, self-regulation, writing efficacy, classroom writing context, average motivation scale]: Band 1 Mean: 0.16, 0.13, 0.13, -0.01, 0.01, -0.29, 0.00, 0.02 Band 1 SD: 0.70, 0.72, 0.44, 0.86, 0.69, 0.67, 0.67, 0.68 Band 2 Mean: 0.00, 0.04, 0.08, -0.09, -0.06, -0.42, 0.00, -0.06 Band 2 SD: 0.80, 0.79, 0.50, 0.94, 0.76, 0.70, 0.70, 0.74 Band 3 Mean: 0.04, 0.05, 0.11, -0.26, -0.17, -0.63, -0.07, -0.13

		Band 3 SD: 0.77, 0.76, 0.49, 0.93, 0.73, 0.70, 0.72, -0.73 Male mean: 0.00, 0.02, 0.07, -0.20, -0.14, -0.48, -0.07, -0.11 Male SD: 0.83, 0.82, 0.51, 0.96, 0.77, 0.74, 0.72, 0.77 Female mean: 0.13, 0.13, 0.14, 0.04, -0.01, -0.42, 0.02, 0.00 Female SD: 0.68, 0.68, 0.43, 0.86, 0.67, 0.67, 0.66, 0.67 Secondary 1 mean: 0.08, 0.09, 0.10, -0.05, -0.01, -0.34, 0.07, -0.01 Secondary 1 SD: 0.81, 0.82, 0.54, 1.01, 0.81, 0.77, 0.74, 0.78 Secondary 3 mean: 0.06, 0.07, 0.10, -0.12, -0.06, -0.42, -0.04, -0.06 Secondary 3 SD: 0.76, 0.75, 0.47, 0.91, 0.73, 0.70, 0.70, 0.72 Secondary 5 mean: 0.06, 0.07, 0.11, -0.17, -0.13, -0.55, -0.08, -0.10 Secondary 5 SD: 0.73, 0.71, 0.43, 0.85, 0.65, 0.64, 0.65, 0.67 The following is my own calculation: (1) Overall SD of population = 0.724
	Effect size	The following is my own calculation: (1) ES for population= -0.049 (2) Standard error of ES estimate = 0.019
	Limitations	Limitations identified by the authors: (1) The study only used quantitative methods; (2) The study only drew upon three secondary schools in Hong Kong.

		Limitation identified by me: (1) In the findings section, the scale of data (+-1/+-2) is unknown. (2) L2 proficiency was measured solely based on the school banding of participants. While school banding and L2 proficiency might be correlated, school banding is not an accurate and reliable measure of one's L2 proficiency. It would be more sensible if the authors simply renamed ""L2 proficiency"" as ""school banding,"" as they are essentially examining the correlations between motivation and school banding, not proficiency.
	Conclusion	(1) Participants of the study were generally not motivated to write in English; (2) There was a significant difference among students from the three bandings in their L2 writing motivation (p,0.001, n^2=0.077); (3) Girls were significantly more motivated than boys in L2 writing (p<0.001, n^2=0.017), although the n^2 values were smaller than 0.01; (4) Secondary 1 students were significantly more motivated than Secondary 3 and 5 students in L2 writing (F=10.302, p<0.001, n^2=0.015).
Comments	Additional comments	N/A

Category	Item	Description
General Information	Date form completed (yyyy/mm/dd)	2024-05-17
	Reference citation	Lee, Kwok Wai. 2018. "Reading Motivation in Chinese and English in Hong Kong Secondary Schools Students: Influence of Schools and Demographic Factors." EdD Thesis, Hong Kong: The Education University of Hong Kong. https://repository.eduhk.hk/en/publications/reading-motivation-in-chinese-and-english-in-hong-kong-secondary-.
	Year of publication	2018
	Gender of first author	Male.
	Database/information source	EdUHK Research Repository.
	Publication type	EdD thesis.
	Language	English.
	Source of funding	N/A
Setting	Context of study	Participants came from seven secondary schools in Hong Kong Island, Kowloon, and the New Territories. The schools covered Bands One, Two and Three, and were all government-aided.
Theoretical Framework	Motivational theory applied (if applicable)	Ryan and Deci's (2000) self-determination theory.
	Motivational components examined	(1) Autonomous English reading motivation; (2) Controlled English reading motivation. Other components irrelevant to this study: (3) autonomous Chinese reading motivation; (4) controlled Chinese reading motivation; (5) teacher support; (6) peer support.

Methods	Research questions	(1) How are the school factors, namely MOI, teacher support and peer support, related to L1 and L2 reading motivation? (2) How are the demographic factors, namely SES, gender and grade, related to L1 and L2 reading motivation? (3) Are there differential patterns of prediction of school factors and demographic factors for L1 and L2 reading motivation?
	Study design	Quantitative.
		Cross-sectional.
		Synchronic study using a 5-point questionnaire combining the Self-Regulation Reading Questionnaire (to measure Chinese and English reading motivations) and Facilitating Conditions Questionnaire (to measure teacher support and peer support).
	Intervention (if applicable)	N/A
	Comparator (if applicable)	N/A
	Measurement tool(s)	To measure English reading motivation, the author adopted the Self-Regulation Questionnaire of Self-Determination Theory questionnaire. There are 12 items measuring autonomous reading motivation, and 12 items measuring controlled reading motivation. Controlled reading motivation is related to reading activities in the schooling/academic context, while autonomous reading motivation is related to reading activities in the recreational context.
		The Self-Regulation Questionnaire of Self-Determination Theory was validated by previous authors (De Naeghel et al. 2012). In the current study,

		the Cronbach's alphas for each parameter also ranged between 0.83 to 0.95, showing an excellend internal consistency. The reading motivation scales were also validated by CFA.
		N/A
Participants	Recruitment method	Formal invitation through mail. The researcher reached out to teachers, and students were assured that participation is voluntary. Before the study began, consent was also collected from principals, parents, and participants.
	Gender composition	426 males and 515 females.
	Demographics/ Eligibility criteria	The participants were from 37 classes, including 12 Secondary 1 classes, 12 Secondary 4 classes, and 13 Secondary 6 classes.
	Sample size	941 participants (426 males and 515 females)
	Attrition	7 participants (5 males and 2 females) out of 948 (0.74%) failed to complete the whole questionnaire. They were hence deleted from the data set.
Outcomes	ESL skill(s) assessed	Students' motivation for English reading was assessed. The aforementioned questionnaire was used.
	Aspect(s) of motivation assessed	(1) Controlled English reading motivation; (2) Autonomous English reading motivation. Irrelevant outcomes: (3) Controlled Chinese reading motivation; (4) Autonomous Chinese reading motivation.
	Outcomes with scores	(1) Overall autonomous English reading motivation (n=941, M=3.04, SD=0.87, a=0.95); (2) Overall controlled English reading motivation (n=941, M=2.50, SD=0.88, a=0.89); (3) Autonomous English reading motivation for males (n=426, M=3.01);

		(4) Controlled English reading motivation for males (n=426, M=2.52); (5) Autonomous English reading motivation for females (n=515, M=3.06); (6) Controlled English reading motivation for females (n=515, M=2.49); (7) Autonomous English reading motivation for CMI students (M=2.82); (8) Controlled English reading motivation for CMI students (M=2.41); (9) Autonomous English reading motivation for EMI students (M=3.16); (10) Controlled English reading motivation for EMI students (M=2.56); (11) Autonomous English reading motivation for Band 1 students (M=3.16); (12) Controlled English reading motivation for Band 1 students (M=2.56); (13) Autonomous English reading motivation for Band 2 students (M=2.91); (14) Controlled English reading motivation for Band 1 students (M=2.33); (15) Autonomous English reading motivation for Band 2 students (M=2.69); (16) Controlled English reading motivation for Band 1 students (M=3.28). The following is my own calculation: (1) Average reading motivation for males (n=426, M=2.765); (2) Average reading motivation for females (n=515, M=2.775); (3) Overall SD of population = 0.916

	Effect size	The following is my own calculation: (1) ES for average reading motivation= -0.011; (2) ES for autonomous reading motivation= -0.057; (3) ES for controlled reading motivation= 0.034. (4) Standard error of ES estimate=0.030
	Limitations	Limitations identified by the author: (1) Since a self-report questionnaire was employed, the findings may be limited. Moreover, participants may have filled in wrong answers, and the fixed choices lacked flexibility. (2) The study did not include outcome variables, such as reading comprehension skills, reading strategies, and reading amount. (3) A generic tool was used to measure teacher support and peer support. Hence, their support in the specific field of reading motivation was not directly measured. Limitations identified by me: (1) The author did not report the SDs of males, females, MOI, and banding in the motivation scores table. (2) The author did not report the number of participants belonging to each MOI and school banding. (3) The author did not report the motivations scores for different gender in Secondary 1, 4, and 6 respectively.
	Conclusion	(1) School and demographic factors influence the autonomous reading motivation of secondary school students in Hong Kong.

		(2) Teacher support correlated positively with autonomous reading motivation in both Chinese and English.
Comments	Additional comments	N/A

Category	Item	Description
General Information	Date form completed (yyyy/mm/dd)	2024-05-11
	Reference citation	Wong, Ruth. 2010. "Carrot or Stick? An Investigation into Motivation Orientations in Learning English among Hong Kong Chinese Students." *Revista Brasileira de Linguística Aplicada* 10 (1): 71–87. https://doi.org/10.1590/S1984-63982010000100005.
	Year of publication	2010
	Gender of first author	Female.
	Database/information source	ProQuest Social Science Premium Collection.
	Publication type	Journal article.
	Language	English.
	Source of funding	N/A
Setting	Context of study	Two secondary schools in Hong Kong. The medium of instruction and banding were not specified.
Theoretical Framework	Motivational theory applied (if applicable)	Gardner and Lambert's (1972) socio-educational model.
	Motivational components examined	(1) Intrinsic motivation; (2) extrinsic motivation.
Methods	Research questions	(1) Are Hong Kong Chinese students more intrinsically or extrinsically motivated? (2) Are there any motivational differences that appear along gender lines?
	Study design	Mixed-methods.
		Cross-sectional.
		Synchronic study utilizing a 5-point Likert scale questionnaire. A semi-structured interview was also conducted with 20 randomly chosen participants.
	Intervention (if applicable)	N/A
	Comparator	N/A

	(if applicable)	
	Measurement tool(s)	"A 5-point Likert scale questionnaire with 16 items was used. 8 items measured intrinsic motivation (i.e. items 3, 4, 9, 10, 11, 12, 13, 15) and the other 8 measured extrinsic motivation (i.e., items 1, 2, 5, 6, 7, 8, 14, 16).
		The questionnaire was worded in simple English to ensure participants' comprehension. Participants were given 15 minutes to complete the questionnaire. The researcher remained in the room to answer potential questions raised by participants."
		The questionnaire was piloted informally by a group of students from the target population. The reliability coefficients were not reported.
Participants	Recruitment method	Not indicated.
	Gender composition	Not indicated.
	Demographics/ Eligibility criteria	74 male (39 in F4, 35 in F6) and 82 female (40 in F4, 42 in F6) students.
	Sample size	Participants are Hong Kong secondary school students aged between 15 and 19. English is a compulsory subject for them, and all had learned English for 12+ years.
	Attrition	156 participants (74 males and 82 females).
Outcomes	ESL skill(s) assessed	Not indicated.
	Aspect(s) of motivation assessed	(1) Intrinsic motivation (measured by 8 5-point Likert scale items); (2) Extrinsic motivation (measured by 8 5-point Likert scale items).
	Outcomes with scores	(1) Intrinsic motivation for female students (n=82, M=4.34, SD=0.56, a=0.756); (2) Intrinsic motivation for male students (n=74, M=3.42, SD=1.58, a=0.743);

		(3) Extrinsic motivation for female students (n=82, M=4.20, SD=0.42, a=0.822); (4) Extrinsic motivation for male students (n=74, M=4.88, SD=0.11, a=0.867). The following is my own calculation: (5) Overall motivation for female students (n=82, M=4.270, SD=0.500) (6) Overall motivation for male students (n=74, M=4.150, SD=1.337) (7) Overall SD of entire population=0.991 (8) Intrinsic SD of entire population=1.249 (9) Extrinsic SD of entire population=0.462 (10) Overall motivation for all students=4.213
	Effect size	The following is my own calculation: (1) ES for intrinsic motivation: -0.737 (2) ES for extrinsic motivation: 1.471 (3) ES for overall motivation: -0.121 (11) Standard Error of ES estimate=0.079
	Limitations	Limitations identified by the author: (1) The study was cross-disciplinary. (2) The sample size was too small. Limitations identified by me: (1) The study failed to measure the participants' current self. (2) Although the questionnaire was met with satisfactory Cronbach's alpha values, the author could have done a more extensive pilot study and reported the reliability coefficients. The "revisions" made after the pilot study remain unknown.

		(3) The study did not indicate the medium of instruction of the participants' schools, which may critically affect their self-concepts and motivation in learning English. (4) The study falsely interpreted Gardner's socio-educational model as an integrative-instrumental dichotomy. It did not measure any other constructs mentioned in the model.
	Conclusion	(1) Overall, Hong Kong secondary school students are more extrinsically motivated in L2 learning. (2) Females have stronger intrinsic motivation than their male counterparts.
Comments	Additional comments	N/A

Category	Item	Description
General Information	Date form completed (yyyy/mm/dd)	2024-05-16
	Reference citation	Yung, Kevin Wai-Ho. 2019. “Exploring the L2 Selves of Senior Secondary Students in English Private Tutoring in Hong Kong.” *System* 80 (February):120–33. https://doi.org/10.1016/j.system.2018.11.003.
	Year of publication	2019
	Gender of first author	Male.
	Database/information source	ProQuest Social Science Premium Collection.
	Publication type	Journal article.
	Language	English.
	Source of funding	N/A
Setting	Context of study	The 18 participants came from 16 schools. Half of them studied in EMI secondary schools, and the other half in CMI schools.
Theoretical Framework	Motivational theory applied (if applicable)	Dörnyei's (2005) L2 Motivational Self-system.
	Motivational components examined	(1) Internal motivation; (2) external motivation.
Methods	Research questions	Based on their English learning experiences, what L2 selves do senior secondary students possess when they subscribe to English private tutoring?
	Study design	Mixed-methods.
		Cross-sectional.
		Synchronic study using a yes/no questionnaire with 14 standard items, and 1 item for participants to provide supplementary answers.
	Intervention (if applicable)	N/A
	Comparator	N/A

	(if applicable)	
	Measurement tool(s)	A yes/no questionnaire was created to elicit participants' attitudes towards learning English and enrolling in English private tutoring courses. In the section evaluating students' attitudes towards English learning, there was an overarching question "I learn English because..." and there are 14 possible reasons to choose from. Participants may tick the ones that apply to them. For item 15, participants may also provide additional answers, although only one participant did so. For the 14 standard options, 9 of them (items 1, 3, 4, 5, 8, 9, 10, 11, 12) measured internal motivation, and 9 of them (items 2, 3, 6, 7, 8, 10, 12, 13, 14) measured external motivation. Items 3, 8, 10, and 12 measure both internal and external motivations.
		The questionnaire's reliability and validity was not tested.
		The researcher interviewed 18 participants, 4 tutors, 13 parents, 4 school teachers, and conducted 12 classroom observations at the tutorial centre. Questions include: (1) Follow-up on questionnaire answers; (2) Can you talk about your background, e.g. your school, family, social life? (3) Can you describe your English learning experience in your life, both in formal schooling and private tutoring? (4) Why is studying English important to you? What is your goal?

		(5) How do you think English should be taught and learned? How have your school and schoolteachers influence the way you learn English? (6) What do you think about the teaching and learning of English in the current education system? (7) How can English learning be more motivating and effective to you? (8) How do you see yourself using English in the future? (9) Why do you attend English tutorial class? (10) Why do you choose star-tutoring? How do you choose it? Advertisements? (11) Why do you choose a live/video class? What are the pros and cons of this mode of tutoring to you? (12) Does your English schoolteacher know that you are attending English tutoring? If so, what is his/her attitude about it? (13) What is your parents' attitude about your attending English tutoring? (14) What do you expect to get from the tutor? How do you expect what you will get from the tutor is similar to or different from what you will get from your schoolteacher? (15) What else can you think of which may help me to understand your English learning experience better?
Participants	Recruitment method	Maximum variation sampling method. 20 participants were selected from 2200 respondents to represent diverse backgrounds and English proficiencies. At the end, 18 participants completed the study.
	Gender composition	9 males and 9 females.

	Demographics/ Eligibility criteria	All participants speak Cantonese as their L1, and English as an L2. They have learned English at school since the age of three, and studied in local secondary schools under the Hong Kong Diploma of Secondary Education Examinations (HKDSE) curriculum. Participants were in Secondary 6.
	Sample size	18 participants (9 males and 9 females).
	Attrition	Not indicated.
Outcomes	ESL skill(s) assessed	N/A
	Aspect(s) of motivation assessed	(1) Internal motivation; (2) external motivation.
	Outcomes with scores	Participant information [name, gender, MOI, HKDSE English language level, proficiency]: Nicola, female, English, 5*, high Zoe, female, Chinese, 5, high Fay, female, Chinese, 5, high Tina, female, English, 4, high Charlene, female, English, 4, high Ken, male, Chinese, 4, medium Luke, male, English, 4, medium Alvin, male, English, 4, medium Diana, female, Chinese, 4, medium Mark, male, Chinese, 4, medium Joanne, female, English, 4, medium Sze, female, English, 3, medium Baron, male, English, 3, low Vincent, male, English, 3, low Peter, male, Chinese, 3, low Wendy, female, Chinese, 2, low Yeung, male, Chinese, 2, low Hong, male, Chinese, 2, low

		For the ease of calculation, and because only one participant answered item 15 (the supplementary answer), only the 14 standard options are taken into account. The number of options (motivators of L2 learning) chosen by each participant are counted below. Nicola: 4 Zoe: 9 Fay: 13 Tina: 10 Charlene: 6 Ken: 11 Luke: 5 Alvin: 2 Diana: 11 Mark: 9 Joanne: 5 Sze: 6 Baron: 4 Vincent: 6 Peter: 5 Wendy: 3 Yeung: 6 Hong: 5 The following is my own calculation: (1) Motivation score for males (n=9, M=5.889, SD=2.667); (2) Mean motivation score for females (n=9, M=7.444, SD=3.432); (3) Standard deviation of population=3.087.
	Effect size	The following is my own calculation: (1) ES for population= -0.504

		(2) Standard Error of ES estimate=0.728
	Limitations	Limitations identified by me: (1) The gender effects found in the study might be due to sampling error. The author selected participants to represent different backgrounds and English proficiencies; however, as a result of the selection, all high proficiency learners were females, and five (out of six) low proficiency learners were males. If one's L2 proficiency and motivation were correlated, this could have possibly led to biased results in the gender gap in motivation. (2) The questionnaire was not validated or piloted. (3) Although the author based the analysis on Dörnyei's (2005) L2MSS , the questionnaire itself was not designed to measured participants' actual self, ideal self, ought-to self, and L2 learning environment. (4) The sample size was rather small. (5) The author's categorization of high, medium, and low proficiency participants was highly subjective. The categorization was solely based on his observation of their writing and speaking competence throughout the study.
	Conclusion	(1) Participants' reasons for learning English were mostly extrinsic; (2) Participants with higher English proficiency tended to learn English with an integrative orientation; (3) Most participants regarded the desire to get a higher grade in public examinations as a key factor of attending English private tutoring classes.
Comments	Additional comments	N/A