



**PEER FEEDBACK EFFECTS ON THE ACCURACY OF PEER
ASSESSMENT OF ESL ARGUMENTATIVE WRITING**

By

XIE XIAO

**Thesis Submitted to the School of Graduate Studies, Universiti Putra
Malaysia, in Fulfilment of the Requirements for the Degree of
Doctor of Philosophy**

April 2024

FBMK 2024 38

All material contained within the thesis, including without limitation text, logos, icons, photographs, and all other artwork, is copyright material of Universiti Putra Malaysia unless otherwise stated. Use may be made of any material contained within the thesis for non-commercial purposes from the copyright holder. Commercial use of material may only be made with the express, prior, written permission of Universiti Putra Malaysia.

Copyright © Universiti Putra Malaysia



Abstract of thesis presented to the Senate of Universiti Putra Malaysia in
fulfilment of the requirement for the degree of Doctor of Philosophy

**PEER FEEDBACK EFFECTS ON THE ACCURACY OF PEER
ASSESSMENT OF ESL ARGUMENTATIVE WRITING**

By

XIE XIAO

April 2024

Chairman : Assoc. Prof. Vahid Nimehchisalem Hossein, PhD
Faculty : Modern Languages and Communication

Evaluative judgement, the ability to assess one's work and others', is crucial for students in their academic and professional journeys. However, current feedback and assessment practices have faced criticism for being unidirectional, overly focused on content and tasks, and passively involving students. To address these concerns, this study proposes reevaluating and redesigning assessment-related activities, especially in terms of the hybrid mode of peer feedback and peer assessment. Peer feedback is a qualitatively ungraded process providing comments and suggestions, and peer assessment primarily focuses on quantitatively assigning scores. While peer feedback's relationship with writing performance has been widely studied, the impact of different roles (feedback providers and receivers) in peer feedback on learners' evaluative judgement about argumentative writing has received little attention. Similarly, peer assessment accuracy in higher education is studied for various reasons, and inaccurate peer assessment ratings may unveil ineffective practices and student bias.

A quasi-experimental design was used to test the impact of different feedback roles on the accuracy of peer assessment. During the five-week intervention, 24 Malaysian undergraduates from a local higher education institution, were assigned into three groups (feedback providers, receivers, and outsiders). A comparison was made between the assessments given by three expert raters and peer raters using the Rasch Partial Credit Model (PCM) to evaluate the accuracy of peer assessment. In addition, the qualitative results were obtained through thematic analysis, scrutinising participants' reflection after each round of peer assessment, as well as open-ended surveys embedded before and after the longitudinal intervention, which enables the researcher to understand the participants' subjective experiences and perceptions.

The quantitative findings indicate that feedback receivers tended to provide the most accurate scores, closely followed by feedback providers, suggesting that those who receive feedback might be more adept at applying evaluative judgement when assessing their peers' essays. Moreover, the Malaysian students participating in the study displayed a heightened awareness and recognition of the significance of Relevance and Adequacy of Content (RAC) in their argumentative essay writing. Furthermore, the study discovered a positive correlation between the writing quality of essays and the accuracy of peer assessment. Qualitative results show the distinctions between the scores and reflections provided by peer raters and expert raters. Moreover, the pre- and post-intervention survey data were analysed from the perspective of peer feedback orientation, unveiling the participants' self-perception regarding utility, accountability, social awareness, and self-efficacy.

In conclusion, the quantitative findings have tested relevant theories within the context of hybrid peer feedback and peer assessment settings, while the qualitative data have revealed the discrepancies between the peer raters and expert raters, as well as the participants' openness towards the peer feedback and peer assessment. From a practical standpoint, the hybrid mode of peer feedback and peer assessment has the potential to enhance student evaluative judgement, optimise writing experiences, and equip students with valuable skills to become good writers.

Keywords: Argumentative Writing, Evaluative Judgement, Peer Assessment, Peer Feedback.

SDG: GOAL 4: Quality Education

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

**KESAN MAKLUM BALAS RAKAN SEBAYA KE ATAS KETEPATAN
PENILAIAN RAKAN SEBAYA DALAM PENULISAN ARGUMENTATIF
BAHASA INGGERIS SEBAGAI BAHASA KEDUA**

Oleh

XIE XIAO

April 2024

Pengerusi : Profesor Madya Vahid Nimehchisalem Hossein, PhD
Fakulti : Bahasa Moden dan Komunikasi

Penilaian evaluatif, yang merujuk kepada keupayaan untuk membuat penilaian terhadap hasil kerja sendiri dan orang lain, merupakan elemen penting dalam perjalanan akademik dan profesional seseorang pelajar. Walau bagaimanapun, situasi semasa amalan penilaian dan maklum balas berhadapan kritikan kerana sentiasa bersifat satu hala, terlalu memberi tumpuan kepada kandungan dan tugas, serta melibatkan pelajar secara pasif. Bagi mengatasi isu-isu tersebut, kajian ini mencadangkan penilaian dan penaik tarafan semula aktiviti berkaitan penilaian, terutama dalam hal mod hibrid maklum balas rakan sebaya dan penilaian rakan sebaya. Maklum balas rakan sebaya adalah proses tanpa penilaian yang dilaksanakan secara kualitatif seperti memberikan komen dan cadangan, manakala penilaian rakan sebaya tertumpu kepada pemberian skor secara kuantitatif. Walaupun hubungan maklum balas rakan sebaya dengan prestasi penulisan telah dikaji secara meluas, kesan peranan yang berbeza (pemberi dan penerima maklum

balas) dalam maklum balas rakan sebaya terhadap penilaian pelajar ke atas penulisan argumentatif masih lagi kurang mendapat perhatian yang sewajarnya. Sama seperti itu, ada berbagai alasan mengapa ketepatan penilaian rakan sebaya dikaji diperingkat pengajian tinggi, dan penilaian rakan sebaya yang kurang tepat kebiasannya membuka tabir kepada praktis yang kurang efektif atau berat sebelah dikalangan pelajar.

Dalam kajian ini, reka bentuk kuasi-eksperimen telah digunakan untuk menguji kesan peranan maklum balas yang berbeza terhadap ketepatan penilaian rakan sebaya. Sepanjang intervensi yang dilaksanakan selama lima minggu, 24 pelajar sarjana muda Malaysia dari institusi pendidikan tinggi tempatan telah dibahagikan ke dalam tiga kumpulan (pemberi maklum balas, penerima, dan orang luar). Satu perbandingan dilakukan antara penilaian yang diberikan oleh tiga penilai pakar dan penilaian rakan sebaya menggunakan Rasch Partial Credit Model (PCM) untuk menilai ketepatan penilaian rakan sebaya. Selain itu, data kualitatif diperoleh melalui analisis tematik terhadap refleksi peserta selepas setiap pusingan penilaian rakan sebaya, serta kaji selidik terbuka yang disisipkan sebelum dan selepas intervensi berterusan, yang membolehkan penyelidik memahami pengalaman dan persepsi peserta secara subjektif.

Dapatan kuantitatif menunjukkan bahawa penerima maklum balas cenderung memberikan skor yang paling tepat, diikuti dekat oleh pemberi maklum balas. Situasi ini menunjukkan bahawa mereka yang menerima maklum balas mungkin lebih mahir dalam mengaplikasikan penilaian evaluatif ketika menilai

esei rakan sebaya mereka. Selain itu, pelajar Malaysia yang mengambil bahagian dalam kajian ini menunjukkan kesedaran dan pengiktirafan yang meningkat tentang kepentingan Relevansi dan Kecukupan Kandungan (RAC) dalam penulisan esei argumentatif mereka. Selanjutnya, kajian ini menemui korelasi positif antara kualiti penulisan esei dan ketepatan penilaian rakan sebaya. Hasil kualitatif menunjukkan perbezaan antara skor dan refleksi yang diberikan oleh penilai rakan sebaya dan penilai pakar. Selain itu, data kaji selidik pra-dan pasca-intervensi dianalisis dari perspektif orientasi maklum balas rakan sebaya, menyingkap persepsi diri peserta mengenai kegunaan, akauntabiliti, kesedaran sosial, dan efikasi diri.

Sebagai kesimpulan, dapatan kajian ini juga telah menguji teori-teori yang relevan dalam konteks maklum balas rakan sebaya hibrid dan tetapan penilaian rakan sebaya secara kuantitatif, sementara data kualitatif telah mendedahkan perbezaan antara penilai rakan sebaya dan penilai pakar, serta keterbukaan peserta terhadap maklum balas rakan sebaya dan penilaian rakan sebaya. Dari sudut praktikal, mod hibrid maklum balas rakan sebaya dan penilaian rakan sebaya mempunyai potensi untuk membina penilaian evaluatif pelajar, mengoptimumkan pengalaman menulis dan melengkapkan pelajar dengan kemahiran yang sangat berharga untuk menjadi seorang penulis yang baik.

Kata Kunci: Penulisan Argumentatif, Penilaian Evaluative, Penilaian Rakan Sebaya, Maklum Balas Rakan Sebaya.

SDG: MATLAMAT 4: Pendidikan Berkualiti

ACKNOWLEDGEMENTS

Completing my doctoral studies is my greatest aspiration, and I have devoted a whole decade of unwavering effort towards this dream. For me, it signifies not only a prestigious academic achievement but, more importantly, a transformative force that has allowed me to become a better version of myself. The journey of completing this doctoral thesis has been the most fulfilling experience, as it demanded a level of rigorous innovation that challenged the depths of my potential. I wish to express my heartfelt gratitude to my three supervisors for their love and guidance. I also want to extend my thanks to my parents for their unwavering support and understanding. I understand that this is not the culmination of my academic journey; rather, it marks a solid starting point, and I am committed to cherishing the adventures that lie ahead.

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Vahid Nimehchisalem Hossein, PhD

Associate Professor
Faculty of Modern Languages and Communication
Universiti Putra Malaysia
(Chairman)

Yong Mei Fung, PhD

Associate Professor
Faculty of Modern Languages and Communication
Universiti Putra Malaysia
(Member)

Yap Ngee Thai, PhD

Associate Professor
Faculty of Modern Languages and Communication
Universiti Putra Malaysia
(Member)

ZALILAH MOHD SHARIFF, PhD

Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 12 September 2024

TABLE OF CONTENTS

	Page
ABSTRACT	i
ABSTRAK	iv
ACKNOWLEDGEMENT	vii
APPROVAL	viii
DECLARATION	x
LIST OF TABLES	xv
LIST OF FIGURES	xvi
LIST OF APPENDICES	xvii
LIST OF ABBREVIATIONS	xviii
CHAPTER	
1 INTRODUCTION	1
1.1 Introduction	1
1.2 Background of the Study	1
1.3 Statement of the Problem	9
1.4 Research Aim	13
1.5 Objectives of the Study	14
1.6 Research Questions	16
1.7 Research Hypotheses	17
1.8 Theoretical Framework	18
1.8.1 The Development of Evaluative Judgement	18
1.8.2 Rater Accuracy Models	21
1.8.3 Peer Feedback Orientation	24
1.9 Conceptual Framework	26
1.10 Definition of Key Terms	29
1.11 Chapter Summary	33
2 LITERATURE REVIEW	34
2.1 Introduction	34
2.2 Developing Evaluative Judgement	34
2.2.1 Connotation of Evaluative Judgement	35
2.2.2 Context of Evaluative Judgement	41
2.2.3 Practices for Developing Evaluative Judgement	47
2.3 The Effect of Peer Feedback Roles on Evaluative Judgement	50
2.3.1 Providing Peer Feedback	50
2.3.2 Receiving Peer Feedback	54
2.3.3 Comparing Providing Feedback and Receiving Feedback	58
2.4 Peer Feedback Orientation	63
2.4.1 Feedback Orientation	63
2.4.2 Dimensions of Peer Feedback Orientation	67
2.5 Peer Assessment Accuracy	73

2.5.1	Traditional Approaches to Examining Peer Assessment	74
2.5.2	Traditional Approaches to Examining Rater Accuracy	77
2.5.3	Rasch Measurement of Peer Assessment Inaccuracy	79
2.5.4	Rasch Modelling of Rater Accuracy	86
2.6	Chapter Summary	95
3	METHODOLOGY	97
3.1	Introduction	97
3.2	Research Design	97
3.3	Location of the Study	99
3.4	Sampling Procedure	100
3.5	Training Procedure	103
3.5.1	Activities during the Training Sessions	110
3.5.2	Activities outside the Training Sessions	113
3.6	Data Collection	118
3.6.1	Instrumentation	118
3.6.2	Data Collection Procedure	122
3.7	Data Analysis	131
3.7.1	Quantitative Data Analysis	132
3.7.2	Qualitative Data Analysis	139
3.8	Validity and Reliability	149
3.8.1	Validity of Quantitative Analysis	149
3.8.2	Reliability of Quantitative Analysis	156
3.8.3	Validity of Qualitative Analysis	162
3.8.4	Reliability of Qualitative Analysis	165
3.8.5	Strategies to Minimise Validity Threats to the Mixed-Methods Design	167
3.9	Chapter Summary	169
4	RESULTS AND DISCUSSION	170
4.1	Introduction	170
4.2	Quantitative Findings and Discussion	170
4.2.1	Descriptive Statistics of Pre- and Post- Intervention Argumentative Writing	171
4.2.2	Summary Statistics for Rasch Partial Credit Modelling	174
4.2.3	Graphic Output of Rasch Partial Credit Modelling	178
4.2.4	Different Peer Feedback Roles on the Accuracy of Peer Assessment	183
4.2.5	Different Rating Domains on the Accuracy of Peer Assessment	188
4.2.6	Different Writing Essays on the Accuracy of Peer Assessment	194
4.3	Comparison of Rating Reflections of Expert Raters and Peer Raters	200

4.3.1	Comparison of Relevance and Adequacy of Content Domain	202
4.3.2	Comparison of Compositional Organisation Domain	205
4.3.3	Comparison of Cohesion Domain	208
4.3.4	Comparison of Vocabulary Domain	212
4.3.5	Comparison of Grammar Domain	215
4.4	Participants' Openness towards Peer Feedback and Peer Assessment	221
4.4.1	Previous Experiences and Challenges	224
4.4.2	Presumed Peer Feedback Benefits	230
4.4.3	Peer Feedback Experiences	235
4.4.4	Peer Assessment Experiences	245
4.5	Chapter Summary	251
5	SUMMARY, CONCLUSION AND RECOMMENDATIONS FOR FUTURE RESEARCH	252
5.1	Introduction	252
5.2	Summary of Major Findings	252
5.3	Implications	255
5.3.1	Implication for Language Learners	255
5.3.2	Implication for Language Educators	257
5.3.3	Implication for Policy Makers	259
5.4	Contribution of the Study	260
5.5	Limitation of the Study	265
5.6	Suggestions for Future Research	268
5.7	Concluding Remarks	273
	REFERENCES	275
	APPENDICES	293
	BIODATA OF STUDENT	343
	LIST OF PUBLICATIONS	344

LIST OF TABLES

Table	Page
2.1 Dimensions of the Peer Feedback Orientation and Feedback Orientation	69
2.2 Sample Ratings from Two Raters	78
3.1 Summary of the Training Schedule in the Experimental and Control Groups	108
3.2 Statistical Indices and Graphical Displays of Rating Accuracy	138
3.3 Themes and Subthemes of Peer Feedback Orientation	144
3.4 Summary of Data Collection and Data Analysis	148
4.1 Descriptive Statistics of Pre- and Post-intervention Argumentative Writing Performance in FPG, FRG and FOG	173
4.2 Summary statistics for Model 1 and Model 2	178
4.3 Calibration of the Feedback Role Facet by Model 1	185
4.4 Calibration of the Peer Rater Facet by Model 1	188
4.5 Calibration of the Rating Domain Facet by Model 1	190
4.6 Calibration of the Writing Essay Facet by Model 1	197
4.7 Ratings on Essays for Relevance and Adequacy of Content Domain	203
4.8 Ratings on Essays for Compositional Organisation Domain	206
4.9 Ratings on Essays for Cohesion Domain	209
4.10 Ratings on Essays for Vocabulary Domain	213
4.11 Ratings on Essays for Grammar Domain	216

LIST OF FIGURES

Figure	Page
1.1 Lens Model II	24
1.2 The Conceptual Framework of the Study	28
3.1 Schematic Illustration of the Process-Genre Instructional Framework	106
3.2 Summary of Data Collection Methods and Instruments	119
3.3 The Flowchart for the General Data Collection Procedure of the Study	123
3.4 The Flowchart for the Data Collection Procedure of Formal Treatment	131
3.5 Mixed-Methods Approach of Data Collection and Data Analysis	147
4.1 Wright Map of Model 1	183
4.2 Correlational Pattern for Essay Performance and Accuracy of Peer Assessment	199

LIST OF APPENDICES

Appendix	Page
A Analytical Rating Scale	293
B Writing Task Prompts	296
C Exemplars for In-Class Group Discussion	297
D Exemplar for Peer Feedback	299
E Article Exemplar for Expert/Peer Assessment	300
F Exemplar for Expert/Peer Assessment Reflection	301
G Average Expert Ratings of 18 Selected Essays as the Criterion Scores	302
H Rating and Frequency of Comments for RAC Domain	303
I Rating and Frequency of Comments for CO Domain	311
J Rating and Frequency of Comments for C Domain	319
K Rating and Frequency of Comments for V Domain	327
L Rating and Frequency of Comments for G Domain	334
M The Pre- and Post-Intervention Survey	342

LIST OF ABBREVIATIONS

FPG	Feedback Provider Group
FRG	Feedback Receiver Group
FOG	Feedback Outsider Group
RAC	Relevance and Adequacy of Content
CO	Compositional Organisation
C	Cohesion
V	Vocabulary
G	Grammar

CHAPTER 1

INTRODUCTION

1.1 Introduction

This chapter initiates with the study's context and then progresses to present the problem statement, which underlines both the real-world issues and the gaps in the existing research concerning peer feedback and peer assessment, particularly emphasising the Malaysian tertiary context. Following this, the research objectives, research questions, and the study's theoretical framework are outlined. This framework includes the evaluative judgement development, rater accuracy models and peer feedback orientation. The chapter wraps up with a discussion about the conceptual framework and the clarification of key terms.

1.2 Background of the Study

In the current landscape marked by the emergence of Artificial Intelligence (AI), our educational methods and assessment strategies are confronting unprecedented challenges. The rapid evolution of AI-driven writing tools, such as ChatGPT, Bing Chat, Jasper.ai, Grammarly, and iWrite, has yielded promising results in terms of students' perceptions of AI-generated instant corrective feedback and its impact on the development of their writing skills (Anders, 2023; Dong & Shi, 2021; Koltovskaia, 2020). Nevertheless, the integration of cutting-edge AI systems, notably ChatGPT, for generating academic essays within educational institutions has ignited concerns among academics, because of "confident responses that seemed faithful and non-

sensical when viewed in light of the common knowledge in these area” (Alkaissi & McFarlane, 2023, p. 3).

A prevailing worry centres around the possible erosion of a vital educational endeavour – the cultivation of evaluative judgement. AI generating a response doesn’t mean it is necessarily accurate, relevant, or of high quality. A person still needs to critically assess the information provided by the AI and determine if it is reliable and useful. This is especially important in fields such as academia, where accuracy and credibility are crucial. As emphasised by Bearman et al. (2024, p. 11), “AI has widened the gap between our capability to produce work, and our capability to evaluate the quality of that work. It is easier than ever to produce something that has the superficial appearance of quality—but knowing if it is good enough for any given purpose requires expertise”. This concern stands out as one of the most significant issues in contemporary education. As Generative Artificial Intelligence continues to reshape our educational horizons, the cultivation of evaluative judgement emerges as an indispensable goal, ensuring that learners thrive not just in the classroom but in the ever-evolving landscape of knowledge acquisition and creation.

Evaluative judgement, the cognitive skill empowers individuals to not only gauge the quality of their own work but also to critically assess the work of others (Ajjawi et al., 2018; Braine, 2003; Tai et al., 2018). Its significance resonates not just within the confines of a singular course but extends to the realms of lifelong learning (Boud & Soler, 2016). Effective evaluative

judgement necessitates a dynamic interplay among individuals, their peers, and established benchmarks of performance. Nevertheless, prevailing assessment and feedback practices frequently fall short in adequately addressing this vital aspect. Certain approaches tend to relegate students to passive roles, positioning them as mere recipients of feedback (Carless et al., 2011; Filius et al., 2018; Forsythe & Johnson, 2017). This passivity hampers the development of their evaluative judgement, leaving them reliant on external appraisals without the capability to discern pertinent criteria for assessment (Gielen et al., 2010). As mentioned by Vasu et al. (2016), taking into account the educational setting of Malaysia, in which teachers are frequently anticipated to be accountable for the academic progress of their students, “peer feedback, although highly valued, turned out to be the least preferred in this context despite the current pedagogical trend that focuses more on students’ active participation in improving their peers’ writing than being dependent on teachers’ feedback” (p. 164).

Recognising the urgency of tackling this concern, recent endeavours have taken strides towards embedding evaluative judgement within diverse assessment-related activities, such as peer feedback and peer assessment. This evolving landscape compels educators and institutions to reconfigure their educational paradigms, fostering an environment where students become discerning evaluators, equipped with the cognitive tools to independently assess their own work and the work of their peers.

Within the domain of educational research, the terms “peer feedback” and “peer assessment” have been occasionally employed interchangeably, a phenomenon underscored by scholars like Gielen et al. (2010). Nonetheless, a critical imperative emerges to distinctly demarcate these two constructs. Peer assessment takes centre stage in the allocation of grades and entails a quantitative scrutiny of peers’ endeavours. Conversely, peer feedback constitutes an ungraded process primarily centred around furnishing qualitative commentary and recommendations. Esfandiari and Myford (2013) emphasised the constructive dimension of peer feedback, devoid of the formalities entailed in graded evaluations. A deeper exploration by Falchikov (2013) unfurled the essence of peer assessment, explicating that it involves students utilising pre-defined criteria and established benchmarks to scrutinise their peers’ undertakings. From this, they then formulate judgement grounded in these criteria. Thus, while both peer feedback and peer assessment converge on the act of students evaluating the work of their peers, their pivotal differentiators reside in the grading facet. Peer feedback stands qualitative, while peer assessment remains quantitatively oriented.

Peer feedback constitutes a dynamic process in which learners engage in offering and receiving evaluations of their peers’ writing, often through written or verbal means, commonly in pairs or small groups (Yu & Lee, 2016). The core objective of this feedback practice, as underscored by Hyland and Hyland (2006), resides in fostering the autonomy of students as writers. Moreover, the work of Butler and Winne (1995) accentuated the pivotal role of feedback in shaping learners’ goal-setting mechanisms, cognitive strategies, appraisal of

learning outcomes, and overall academic achievements. In the Malaysian tertiary context, Husin and Ariffin's (2008) study delved into the perceptions of students towards the practice of peer feedback within the context of an English as a Second Language (ESL) classroom. The research revealed that students generally welcomed peer feedback as it offered them the advantage of gaining diverse perspectives from a broader range of individuals. This suggests that students recognise the value of multiple viewpoints in enhancing their understanding and learning process. However, the study also uncovered a crucial aspect that could influence the effectiveness of peer feedback. The linguistic competence of the peers providing feedback, their attitudes towards the process, and the cultural values they hold were identified as factors that could potentially affect the quality and credibility of the feedback given. This implies that the ability of the peer to communicate effectively in English, their approach and mindset in offering feedback, and their cultural background could all play a role in shaping the feedback's impact on the recipient. Moreover, peer feedback is not as widely utilised as it could be in the Malaysian context, according to Vasu et al. (2016). This may be due to the fact that Malaysian students are hesitant to fully trust the feedback provided by their peers, preferring instead to rely on teachers' feedback. They may feel that it is the teacher's responsibility to provide feedback and that they do not have the necessary linguistic competence to offer meaningful feedback to their peers' work. In order to encourage the use of peer feedback in Malaysian classrooms, it is important to address these concerns and provide students with the necessary support and training.

While an extensive body of research spanning the last three decades has centred on scrutinising the influence of diverse peer feedback methodologies on holistic writing enhancement (Connor & Asenavage, 1994; Gao et al., 2019), along with investigating the attitudes and encounters of both students and educators towards peer feedback (Chang, 2016), a relatively modest portion of this inquiry has been channelled into comprehending the effects of distinct roles (Allen & Katayama, 2016; Gao et al., 2023; Huisman et al., 2018), particularly those pertaining to feedback providers and receivers, within the sphere of peer feedback. The ensuing exploration specifically probes into how these roles interact to mould the development of evaluative judgement in the context of argumentative writing. Furthermore, it's worth noting that both roles, that of the feedback provider and feedback receiver, come with distinct tasks and responsibilities. The execution of these roles is not uniform and is, in fact, profoundly influenced by the diverse range of individual differences present among students. These differences encompass their beliefs, perceptions, emotional dispositions, and domain knowledge, as elucidated by Alqassab et al. (2019). Despite the paramount importance of these individual nuances, there has been a noticeable dearth of attention afforded to understanding the unique needs and beliefs of students concerning peer feedback. This oversight in research is quite conspicuous in the academic landscape, as highlighted by scholars like Dawson et al. (2019), Mulliner and Tucker (2017), and Srichanyachon (2012). It's crucial to underscore that engaging in the provision and reception of feedback entails a certain degree of openness from individuals, a quality often referred to as feedback orientation. This willingness to embrace and effectively utilise feedback is essential for personal and

academic growth. However, while this trait of openness is undoubtedly vital in the context of peer feedback processes, there has been a noticeable paucity of research addressing the personal factors that influence a student's level of peer feedback orientation (Chen, 2010).

Peer assessment, as meticulously outlined by Topping (1998), entails the appraisal of peers' learning outcomes or products, encompassing considerations such as quantity, level, value, worth, quality, and success. In recent times, the practice of peer assessment has progressively garnered prominence within the realms of language education and higher learning, emerging as a potent tool for multifaceted objectives (Evans, 2013, 2016; Gielen & De Wever, 2015). Its role extends to acquainting students with the diverse expectations of quality held by distinct user groups, thus offering them a comprehensive grasp of the benchmarks and criteria applied to gauge their own work. This immersive involvement in evaluating the work of their peers not only cultivates their awareness of quality parameters but also exposes them to a gamut of examples and varying levels of accomplishment (Fowler, 2007; Hartley et al., 2003; Lee, 2017; Lim, 2013). Despite the potential advantages associated with peer assessment, researchers persistently grapple with concerns regarding the reliability of ratings bestowed by peer raters (Falchikov & Goldfinch, 2000; Li et al., 2016). The findings from Cheah et al. (2018) highlight a cultural aspect of peer assessment within the Malaysian educational context. The belief among undergraduates that teachers possess greater authority and expertise to evaluate academic work is indicative of the respect for the traditional role of educators. This perception

could be influenced by the hierarchical structure often present in educational institutions, where teachers are seen as the ultimate arbiters of knowledge and learning. Moreover, the tendency of students to award higher scores to friends and popular peers suggests that personal relationships and social dynamics can influence the objectivity of peer assessment. This could be due to a desire to maintain positive social ties or to avoid conflict, which may inadvertently compromise the fairness and accuracy of the evaluation process. Therefore, within the domain of higher education, the exploration of peer assessment accuracy has commanded significant scholarly attention, driven by several pivotal factors. Firstly, the aptitude of students to accurately and consistently evaluate their peers bears testament to the successful internalisation and application of the instructor's grading criteria. Secondly, peer assessment serves as an avenue for students to gain deeper insights into and adeptness with scoring rubrics, fostering their comprehension and confidence in their use. Lastly, instances of erroneous peer assessments accentuate the necessity to scrutinise elements such as suboptimal peer assessment methodologies and latent biases among students.

In its entirety, this thesis aims to propel our comprehension of how the amalgamation of peer feedback and peer assessment can effectively elevate the prowess of evaluative judgement within the realm of argumentative writing. The culmination of this research promises to yield invaluable revelations for educators, curriculum designers, and policymakers alike. Armed with these insights, they will be empowered to formulate pedagogical strategies grounded in empirical evidence. These strategies will not only nurture the growth of

evaluative judgement but also instil the bedrock of lifelong learning competencies among students, thereby paving the way for a more adept and discerning generation of learners.

1.3 Statement of the Problem

Despite the wealth of studies on peer feedback and peer assessment, the existing literature lacks comprehensive investigations into the effects of diverse roles within the peer feedback process on learners' evaluative judgement, particularly concerning argumentative writing. In-depth research on the impact of different feedback roles on the accuracy of peer assessment and its hybrid-mode application in teaching practice is of great significance for promoting educational innovation and improving students' learning outcomes. Furthermore, the evaluation of raters' behaviour and the accuracy of peer assessment in scoring essays pose significant challenges in educational research. Traditional approaches, such as inter-rater reliability estimates using Pearson's r coefficient, have limitations that hinder the comprehensive understanding of rater behaviour. Moreover, the reliance on single-point peer assessment in previous research fails to provide stable results or capture developmental trends in rater behaviour and scale utility. Additionally, despite the critical importance of participant engagement and openness to peer feedback and peer assessment, there is a research gap in beliefs and perceptions about providing and receiving peer feedback.

To start with, in Malaysia, students at the tertiary level greatly prefer teacher feedback, as they often believe that teachers have the greatest expertise

(Vasu et al., 2016), while they do not have the necessary linguistic competence to give meaningful feedback on their peers' work. Thus, recent peer feedback practices have been criticised for being unidirectional and overly focused on content, causing students to be passive recipients. In the particular context of ESL writing, there has been extensive research conducted on peer feedback, but it has primarily concentrated on text revisions and writing quality, with significant attention given to the linguistic or textual aspects of student writing (Liu & Yu, 2022; Yu & Lee, 2016), while neglecting higher-order cognitive abilities such as evaluative judgement. Moreover, feedback providers and feedback receivers demonstrate different mechanisms when engaging in peer feedback (Latifi et al., 2021), and the influence of various feedback roles on the accuracy of peer assessment remains largely unexplored. The learning mechanisms involved in providing peer feedback behaviours involve initiating problem detection, facilitating problem diagnosis, and implementing modification strategies. The learning mechanisms associated with receiving peer feedback encompass obtaining information about current performance and desired performance, and devising strategies to close this gap. By exploring the impact of different feedback roles on the accuracy of peer assessment, we can gain a better understanding of how to design and implement peer feedback and peer assessment activities to enhance students' learning outcomes. This may involve investigating the characteristics of different feedback roles, their behaviours in the feedback process, and how these behaviours influence the accuracy of the peer assessment. Designing peer feedback and peer assessment as a hybrid model of pedagogy means combining the two approaches to leverage their synergistic effects to cultivate

evaluative judgement. In this way, students can better understand the criteria and processes of assessment, improve their own assessment abilities, and develop critical thinking and self-reflection habits.

Moreover, when it comes to peer assessment accuracy, it is worth noting that the prevalent approach of utilising inter-rater reliability measures, particularly Pearson's r coefficient, for the scrutiny of rater behaviour comes with inherent constraints. One pivotal concern lies in the fact that the inter-rater reliability coefficient between two raters, A and B, lacks specificity towards each individual rater. Additionally, the application of Pearson's r merely indicates the likeness in scoring performances between raters, disregarding any potential statistically significant disparities that might persist. To surmount these limitations, Rasch models have become prominent as an alternative to conventional correlation analysis when evaluating the accuracy of raters. They offer a robust way to determine the degree of alignment between the scores given by peer raters and those given by expert raters. By doing so, these models not only reveal the absolute match but also provide individual-level statistics for every element of the assessment facets being considered. Yet, it is imperative to acknowledge that the implementation and suitability of these models in assessing the precision of peer ratings of ESL argumentative writing warrant further exploration. Hence, this research sought to bridge this gap by incorporating Engelhard's Partial Credit Model (PCM) within a Rasch-based framework. The primary objective was to evaluate the accuracy of peer assessment in the context of ESL argumentative writing.

In addition, Luo and Chan (2023a) maintained that the ability to make evaluative judgement is not a fixed characteristic that a student either has or lacks; rather, it is a process by which students engage with multiple interconnected facets. This not only supports a more process-oriented approach to developing students' evaluative judgement skills in the curriculum but also emphasizes the need for an integrated curriculum design that takes into account the various aspects emphasized by evaluative judgement and the synergies among them. As a prevailing approach to cultivate evaluative judgement, peer assessment in the current corpus of literature predominantly leans on single-point pedagogical design at the end of the course (Han, 2018; Han & Zhao, 2021). This design poses constraints on the depth of insights garnered concerning the precision of peer assessment, thereby neglecting to capture longitudinal patterns in rater conduct and scale efficacy (Han, 2017). This design limitation inherently hampers researchers from observing and contrasting the performance of rating scales and rater behaviour across varied tasks. To counteract this constraint, as well as to form a substantial conclusion about the efficiency and efficacy of peer assessment practices, this study implemented a three-round peer assessment design by adopting a longitudinal stance, researchers can uncover nuanced trends and dynamics that unfold over time, offering a richer and more nuanced picture of the interplay between raters, tasks, and the assessment scales.

Lastly, Over the past three decades, a vast amount of research has focused on examining the impact of peer feedback on improving writing skills as a whole. However, a relatively small portion of this research has been dedicated

to understanding the differences in students' cognitive, evaluative, and socio-affective processes (Cheng & Zhang, 2024; Liu & Yu, 2022). Moreover, there has been a significant lack of research exploring students' perceptions of the orientation of peer feedback (Kasch et al., 2021, 2023). It is important to note that engaging in the process of providing and receiving feedback requires a certain level of openness from individuals to make the most of peer feedback (Boud & Molloy, 2013; Carless & Boud, 2018; Huisman et al., 2019). The study sought to address this gap in knowledge by employing qualitative data analysis and endeavoured to unearth the underlying reasons behind the challenges encountered when providing constructive feedback and accurately rating specific essays in the context of peer assessment, and to delve into the depths of participants' self-perceptions of their roles as feedback provider, feedback receiver or feedback outsider in peer feedback processes. By shedding light on these aspects, the research aimed to contribute valuable insights into the dynamics of peer feedback and peer assessment of Malaysian university students in higher education context.

In summary, these gaps in knowledge necessitate a more robust and nuanced investigation into the accuracy of peer assessment in ESL argumentative writing, as well as the openness of participating in providing and receiving peer feedback and peer assessment.

1.4 Research Aim

To address the current gaps concerning evaluative judgement, the research aims of this thesis were as follows:

- 1) To assess and compare the impact of different peer feedback roles (feedback provider, feedback receiver or feedback outsider) on the accuracy of peer assessment.
- 2) To evaluate and compare the complexity of accurately peer assessing essays across distinct analytical domains (Relevance and Adequacy of Content, Compositional Organisation, Cohesion, Vocabulary, and Grammar).
- 3) To investigate and compare the difficulties associated with accurately peer assessing argumentative essays of differing proficiency levels.
- 4) To examine and compare the essay features identified by participants during peer assessment with those acknowledged by expert raters.
- 5) To explore participants' openness to peer feedback and peer assessment.

1.5 Objectives of the Study

In order to address the aforementioned research gaps, this thesis aimed to achieve the following objectives:

1) To assess and compare the impact of different peer feedback roles on the accuracy of peer assessment. This objective aimed to understand how different feedback roles (feedback provider, feedback receiver and feedback outsider) influence the accuracy of peer assessment and to identify the most effective feedback roles in improving assessment accuracy.

2) To evaluate and compare the difficulty of accurately peer assessing essays based on different analytical assessment domains. This objective aimed to identify the domains (Relevance and Adequacy of Content, Compositional

Organisation, Cohesion, Vocabulary, and Grammar) that pose challenges in peer assessment and to determine the level of difficulty associated with each criterion, contributing to the understanding of the specific aspects that require attention and improvement in the assessment process.

3) To investigate and compare the challenges of accurately peer assessing argumentative essays of varying levels of proficiency. This involved analysing the impact of higher and lower quality argumentative essays on the accuracy of peer assessment. Through this investigation, the study aimed to contribute to a better understanding of the complexities involved in accurately peer assessing argumentative essays of different levels of proficiency.

4) To investigate and compare the essay features identified by participants in their peer assessment process with those identified by expert raters regarding different writing domains (Relevance and Adequacy of Content, Compositional Organisation, Cohesion, Vocabulary, and Grammar). This objective aimed to explore the agreement and disparities between peer assessments and expert assessments in identifying significant essay features. By comparing the perspectives of participants and experts, this objective sought to provide insights into the accuracy of peer assessment in recognising critical essay elements.

5) To explore the openness of participants towards peer feedback and peer assessment. This objective involved investigating participants' perceptions, beliefs, and confidence levels related to engaging in peer feedback and peer

assessment. Understanding participants' attitudes and self-efficacy would provide valuable insights into their experiences and perspectives, enabling the development of targeted interventions to enhance the effectiveness of peer assessment practices.

1.6 Research Questions

RQ1. Which peer feedback roles (feedback provider, feedback receiver or feedback outsider) influence the accuracy of peer assessment?

RQ2. Which analytical assessment domains (Relevance and Adequacy of Content, Compositional Organisation, Cohesion, Vocabulary, or Grammar) demonstrate higher likelihood of being peer assessed accurately?

RQ3. What is the relationship between the quality of argumentative writing essays and their difficulty/easiness of being peer assessed accurately?

RQ4. What is the difference between expert raters' and participants' assessments of the essays with regard to each domain of argumentative writing (Relevance and Adequacy of Content, Compositional Organisation, Cohesion, Vocabulary, and Grammar)?

RQ5. How do the participants perceive their openness towards peer feedback and peer assessment?

1.7 Research Hypotheses

To appropriately address the research questions, the following research hypotheses were proposed:

Hypothesis 1: Different types of peer feedback roles (feedback provider, feedback receiver or feedback outsider) significantly influence the accuracy of peer assessment.

Hypothesis 2: Specific assessment domains (Relevance and Adequacy of Content, Compositional Organisation, Cohesion, Vocabulary, and Grammar) exhibit different likelihood to be peer assessed accurately.

Hypothesis 3: Essays with higher level of performance would be more likely to be peer assessed accurately compared to those with lower level of performance.

This study proposed several hypotheses to investigate the effects of feedback behaviours, assessment domain, and performance levels on the accuracy of peer assessment in argumentative writing. Hypothesis 1 suggested that certain feedback behaviours would result in higher accuracy compared to others. This hypothesis aimed to identify specific feedback approaches or strategies that are more effective in improving the accuracy of peer assessments. Hypothesis 2 focused on the assessment domain used in peer evaluation. It posited that specific assessment domains, such as Relevance and Adequacy of Content, Compositional Organisation, Cohesion, Vocabulary,

and Grammar, would exhibit higher likelihood of being peer assessed accurately. This hypothesis aimed to identify which assessment domains would contribute most significantly to the accuracy of peer assessments in argumentative writing. Hypothesis 3 addressed the relationship between performance levels in argumentative essays and the accuracy of peer assessment. It was expected that essays with higher level of performance would be more likely to be accurately peer assessed compared to those with lower level of performance. This hypothesis aimed to investigate whether the quality of the essays being assessed would influence the accuracy of peer evaluations. By examining these hypotheses, this study sought to gain insights into the factors that impact the accuracy of peer assessment in argumentative writing, including feedback behaviours, assessment domains, and performance levels.

1.8 Theoretical Framework

1.8.1 The Development of Evaluative Judgement

In their study, Tai et al. (2016) delved into the concept of evaluative judgement and proposed that it consists of two interconnected components that work in harmony. The first component involves grasping the essence of what constitutes quality, while the second component entails applying this understanding to evaluate various works, whether they belong to oneself or others. The authors emphasise that evaluative judgement is not universally applicable but rather depend on specific criteria. As individuals gain expertise in a particular subject or disciplinary area, they acquire the ability to assess

the quality of work within that domain effectively.

Tai et al. (2016), as well as Ajjawi et al. (2018) went on to highlight several existing practices that hold the potential to foster the development of evaluative judgement. These practices encompass self-assessment, peer assessment, feedback, rubrics, and exemplars. By engaging in self-assessment, individuals can reflect on their own work, critically analyse its quality, and identify areas for improvement. Peer assessment provides valuable opportunities for individuals to evaluate the work of their peers, fostering a deeper understanding of quality domains and enhancing one's ability to make evaluative judgement. Feedback, whether from instructors, mentors, or peers, plays a vital role in guiding individuals' evaluative judgement by providing constructive insights and highlighting strengths and weaknesses in a given piece of work. Additionally, the utilisation of rubrics and exemplars can significantly contribute to the development of evaluative judgement. Rubrics establish clear criteria and standards for assessing work, enabling individuals to align their judgement with predetermined benchmarks. Exemplars, on the other hand, serve as concrete examples of high-quality work, allowing individuals to gain a tangible reference point for evaluating the quality of their own work or the work of others.

Several studies have highlighted the multifaceted benefits of peer assessment in the realms of formative education and pedagogy (Clark, 2012; Dochy et al., 1999; Liu & Carless, 2006; Panadero, 2016). Moreover, these studies emphasise the pivotal role of peer assessment in fostering the development

of evaluative judgement (Nicol et al., 2014). By immersing themselves in the realm of established standards and norms and conducting comparisons against them, students gain a heightened comprehension of what constitutes high-quality work (Ajjawi et al., 2018; McConlogue, 2012; Tai et al., 2016). A wide array of peer assessment activities, both formal and informal, hold the potential to significantly contribute to the cultivation of evaluative judgement. When students explicitly concentrate on the aspect of evaluative judgement, they are able to diligently scrutinise the characteristics that define quality within the work of their peers, subsequently transferring this discernment to their own assignments and other academic pursuits. The advantages of engaging in peer assessment extend well beyond the immediate benefits. Students who participate in peer assessment activities acquire invaluable skills that extend into real-world scenarios, including professional environments.

The field of feedback research has made significant strides, moving beyond the simple provision of information to students about their work. Instead, the focus has shifted towards understanding the impact of feedback on students' subsequent work and actively involving them as participants in forward-looking conversations about their learning (Boud and Molloy, 2013; Carless et al., 2011; Merry et al., 2013). In the context of evaluative judgement, the development of feedback places particular emphasis on fostering dialogue. This dialogue serves to clarify the relevant domains and establish standards of quality, enabling students to refine their own judgement about their work. It also helps them formulate actionable steps that arise from their understanding and appreciation of their work, as well as the pertinent information provided by

other sources. When providing feedback, it is not solely focused on the work itself but also on students' own judgement regarding their work. Students require specific information to enhance their evaluative judgement, such as whether they have chosen appropriate criteria, made sound decisions in relation to their work, and what steps they need to take to further develop these skills. Feedback serves as a guiding tool in this process, enabling students to critically reflect on their judgement, make adjustments, and acquire the necessary skills for future improvement.

1.8.2 Rater Accuracy Models

Rater-mediated assessment encompasses assessment systems that rely on manual scoring to effectively evaluate student performance. This approach is prevalent in performance assessments, which offer students opportunities to measure various intricate facets of their learning. Among the diverse forms of performance assessment in educational settings, writing assessment stands out as a widely used method that heavily relies on the ratings provided by raters. To better comprehend and enhance rater-mediated assessment practices, Brunswik's (1952) Lens model offers a valuable conceptual framework, consisting of three essential components: an ecosystem, a judgement system, and a set of cues. Within this model, the ecosystem identifies a distal variable that serves as the criterion, while the study explores the relationship between the distal variable and a set of cues selected as indicators of the ecosystem construct. By comparing the utilisation of cues in ecosystems and judgement systems, the Lens model aims to advance our understanding of rater-mediated assessment practices (Brunswik, 1952;

Cooksey, 1996).

Engelhard (2013) presented two distinct Lens models that offer a conceptual framework for understanding the various constructs involved in rater-mediated assessment. The first Lens model, referred to as Lens Model I in this study, focuses on assessing students' writing ability through observed ratings assigned by human raters (Engelhard, 1992, 1994, 2013; Engelhard et al., 2018).

Lens Model I provides insights into the assessment of writing ability by examining the interplay between students' performance across a range of cues and the empirical judgement made by raters. These cues encompass factors such as rater severity, the interpretation of analytical scoring domains, and the utilisation of predefined rating categories. Rater severity pertains to the consistent tendency of raters to assign higher ratings compared to the standard ratings (Wolfe, 2014). The presence of rater effects can lead to rating inconsistencies and may inaccurately reflect students' actual writing abilities. Consequently, Lens Model I proves valuable in describing the scoring process employed in rater-mediated assessment and investigating the potential influences on rater scoring decisions, including systematic scoring errors and biases.

The Lens Model II in Figure 1.1, as proposed by Engelhard (2013), places a strong emphasis on the assessment of rater accuracy. It aims to evaluate the extent to which raters' observed ratings align with the standard ratings, thus

providing insights into the rater's ability to rate performances accurately. By examining the difference between standard and observed ratings, accuracy ratings offer a means to gauge both the presence and degree of accuracy in the rater's scoring process. This enables us to position a rater's accuracy on a continuum of potential scoring accuracies. Multiple approaches can be adopted to establish standard ratings. One approach involves determining the standard score as the average rating assigned by all raters involved (Wolfe & McVay, 2012), often referred to as the "wisdom of crowds" approach. Alternatively, a group of expert raters can engage in collective discussions to resolve the criterion score (Engelhard, 1996). Engelhard and Wind (2017) defined expert raters as trained individuals possessing extensive knowledge of the assessment system in question. Within Lens Model II, rater accuracy is influenced by various factors unique to each rater, which are empirically identified. These factors encompass elements such as prior scoring experience, perceptions of essay characteristics, and personal preferences regarding the use of scoring categories. These components collectively contribute to understanding and evaluating rater accuracy, a fundamental aspect of Lens Model II.

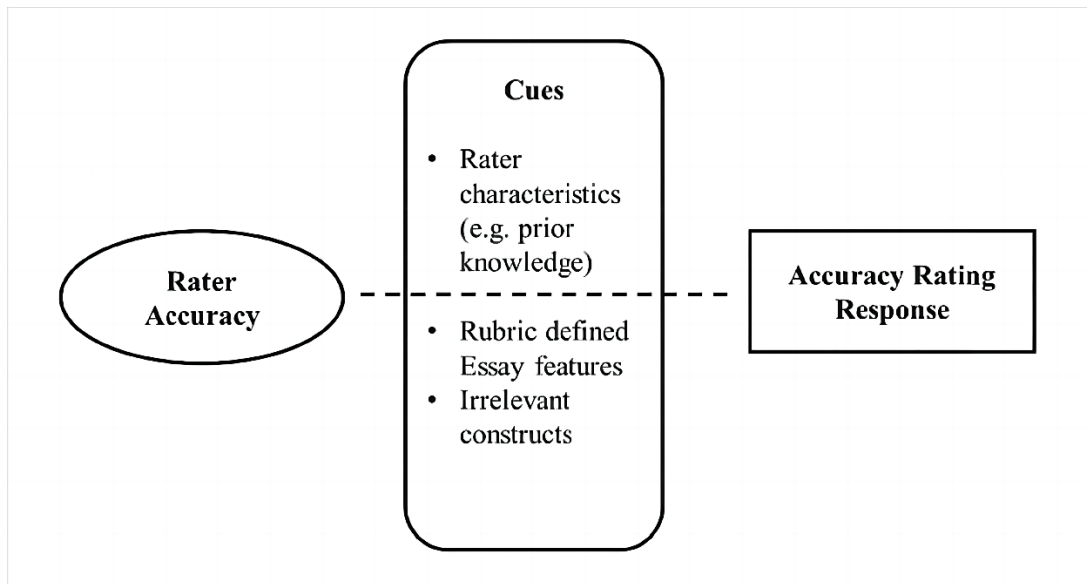


Figure 1.1 : Lens Model II
(Source: Wang & Engelhard, 2019, p. 4)

1.8.3 Peer Feedback Orientation

Engaging in the reciprocal process of providing and receiving peer feedback inherently requires a certain level of openness, often referred to as peer feedback orientation (Kasch et al., 2022, 2023). This attribute plays a pivotal role in shaping students' attitudes and behaviours in the context of peer feedback interactions. Students who exhibit a lower peer feedback orientation, indicating reduced openness towards both giving and receiving peer feedback, are anticipated to hold distinct perspectives, motivations, and requirements compared to their counterparts with a higher peer feedback orientation. This differentiation in peer feedback orientation can significantly impact the dynamics of peer feedback sessions within educational settings. Students with lower peer feedback orientation may approach these interactions with a more cautious or reserved attitude, potentially driven by concerns about evaluation or uncertainty about their feedback-giving abilities. In contrast, those with a higher peer feedback orientation might be more willing to engage openly and

constructively, driven by a stronger belief in the value of peer feedback for their own learning and growth.

The research conducted by Kasch et al. (2023) has brought to light the significance of four key dimensions, namely utility, accountability, social awareness, and self-efficacy, in shaping students' peer feedback orientation. These dimensions have emerged as crucial determinants influencing how students perceive and engage with peer feedback processes within educational settings. When considering their application in the context of educational peer feedback, the qualitative findings (Kasch et al., 2023) revealed that these dimensions took on a slightly altered meaning compared to their interpretation within a workplace environment (Linderbaum & Levy, 2010). Peer feedback in an educational context fundamentally diverges from the experience of receiving feedback in a workplace setting from colleagues. This nuanced shift in meaning underscores the dynamic nature of peer feedback orientation, showcasing how it can be influenced by the specific context in which it is applied.

Additionally, when examining the elements of feedback orientation as delineated by Linderbaum and Levy (2010), it's noteworthy that their focus primarily centred on the orientation of feedback receivers. In contrast, the qualitative investigation conducted by Kasch et al. (2023) expanded its scope to encompass both the act of receiving and providing feedback. This expanded perspective recognises that engaging in these two aspects of the feedback process entails distinct sets of tasks and responsibilities. Crucially, this

broader examination also brings to the fore an essential distinction. It highlights that an individual's propensity and willingness to provide feedback might not necessarily align with their openness to receive feedback. This insight is in line with the observations made by Alqassab et al. (2019), who underscored that these are separate dimensions influenced by individual differences. This distinction is pivotal in the realm of peer feedback within educational contexts. It suggests that educators should not assume that students who are comfortable providing feedback will automatically exhibit the same openness when receiving it.

Understanding this contextual variance is essential for educators and researchers as it allows for a more nuanced examination of providing and receiving peer feedback processes. It emphasises the need to consider the unique educational context when designing and implementing peer feedback interventions. By recognising these contextual subtleties, educators can tailor their approaches to foster a more effective and meaningful peer feedback experience for students, ultimately enhancing their learning outcomes.

1.9 Conceptual Framework

There exist several established practices that hold the potential to foster the development of evaluative judgement, notably peer feedback and peer assessment (Ajjawi et al., 2018; Tai et al., 2016). This study embarks on an exploration into the intricate mechanisms at play within the evaluative judgement construct, particularly focusing on the mediation of rater accuracy - a distal, unobservable variable - through a series of cues, ultimately

manifesting as peer assessment accuracy. As illustrated in Figure 1.2, the expansive peripheral circle encapsulates the ecosystem of evaluative judgement. Within this framework, the smaller circle on the left signifies a latent variable, specifically denoted as rater accuracy. In contrast, the square on the right denotes observed responses, designated as peer assessment accuracy. Situated at the core of the overarching circle, we find the cues, initially characterised as proximal peripheral cues by Brunswik (1955). These cues serve as the mediators in the process of gauging the construct. This intricate web of interrelationships underscores the multifaceted nature of evaluative judgement. It unveils how seemingly imperceptible factors, embodied within the concept of rater accuracy, are translated into tangible and measurable outcomes in the form of peer assessment accuracy. The cues, residing at the heart of this mechanism, act as conduits through which the latent construct of rater accuracy is observed and quantified.

Rooted in Engelhard's (2013) Lens Models of Rater-Mediated Assessment, this study delves into three key factors that impact the accuracy of peer assessment. These factors are the distinct roles undertaken by peers in providing and receiving feedback, including feedback providers, receivers, and outsiders; the diverse writing domains, encompassing Relevance and Adequacy of Content (RAC), Compositional Organisation (CO), Cohesion (C), Vocabulary (V), and Grammar (G); and the proficiency levels of essays, ranging from well-written to poorly-written pieces.

The central objective of this thesis is to thoroughly examine the effectiveness of practices that encompass peer feedback and peer assessment in enhancing students' evaluative judgement in ESL argumentative writing. Since there is a dearth of empirical research investigating evaluative judgement, this study endeavours to bridge this gap by implementing an extensive five-week writing training programme.

Across the span of the five-week training programme, participants are immersed in various writing tasks that encompass giving and receiving peer feedback, followed by the process of peer assessment facilitated by analytical rubrics. This integration of quantitative peer assessment scores with qualitative rating reflections offers a holistic approach to both evaluating and understanding the enhancement of evaluative judgement in ESL argumentative writing.

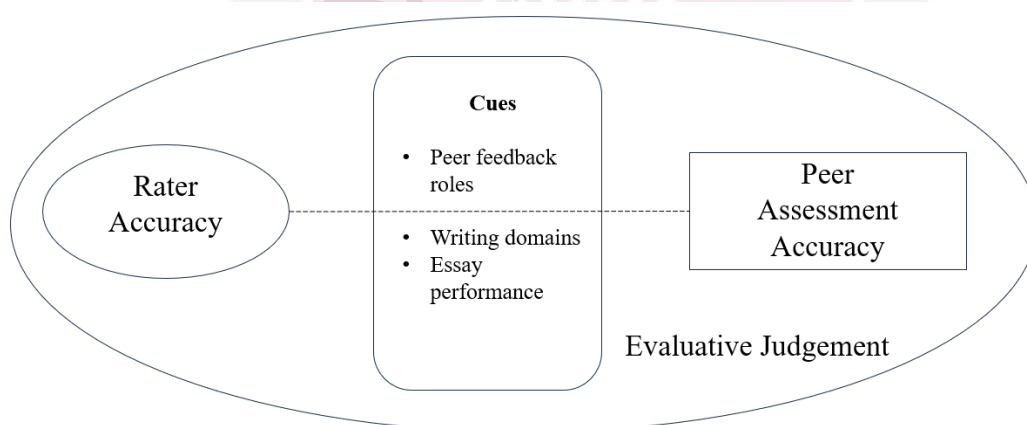


Figure 1.2 : The Conceptual Framework of the Study

1.10 Definition of Key Terms

1) Evaluative judgement: The concept of evaluative judgement encompasses the capacity to evaluate and make well-informed decisions regarding the quality and value of work, be it one's own or that of others (Ajjawi et al., 2018; Tai et al., 2018). It entails engaging in a process of critical analysis, whereby multiple factors and domains are carefully examined and considered to assess the merit, effectiveness, or overall worthiness of a specific piece of work. Evaluative judgement involves going beyond surface-level observations and delving into deeper layers of analysis, allowing individuals to form thoughtful and reasoned assessments. By cultivating this ability, individuals gain a more comprehensive understanding of the strengths, weaknesses, and overall significance of various works, enabling them to make informed choices and contribute to ongoing improvement and advancement. Developing the skill of evaluative judgement empowers individuals to make informed decisions based on a thorough evaluation of pertinent factors, leading to more thoughtful and effective contributions within their respective fields.

2) Peer feedback: Peer feedback is an interactive process in which learners provide and receive feedback on each other's writing, either in written or oral form, typically in pairs or small groups (Yu & Lee, 2016). It involves sharing opinions, suggestions, and constructive criticism aimed at improving the quality and effectiveness of the written work. Peer feedback, in contrast to peer assessment, does not involve assigning grades and is primarily focused on providing qualitative insights. As highlighted by Esfandiari and Myford (2013), peer feedback is characterised by its ungraded nature and emphasises the

provision of constructive comments, suggestions, and observations. This qualitative approach allows students to offer personalised and detailed feedback to their peers, focusing on the strengths and areas for improvement in their work. Without the constraints of assigning grades, peer feedback encourages a supportive and collaborative learning environment where students can engage in thoughtful discussions and provide valuable insights to enhance the quality of their peers' work. By emphasising qualitative aspects, peer feedback fosters the development of critical thinking, communication skills, and a deeper understanding of the subject matter among students.

3) Peer assessment: Peer assessment is a method in which individuals, typically of similar status or within the same learning group, evaluate and judge the products or outcomes of their peers' learning. It involves considering various aspects such as quantity, level, value, worth, quality, or success of the work produced by peers (Topping, 1998). Peer assessment encourages active engagement and fosters a deeper understanding of the subject matter through the evaluation process. In the context of evaluating student work, peer assessment assumes a graded and quantitative nature. As aptly described by Boud and Falchikov (2005), peer assessment entails students employing predefined criteria and established standards to scrutinise the work of their fellow peers, enabling them to form judgement about the quality of the work. This process involves applying objective measurements and grading mechanisms to assess various aspects of their peers' submissions, ensuring a systematic and quantitative evaluation. In essence, peer assessment provides students with the opportunity to critically analyse their peers' work

based on specific criteria and standards, contributing to a comprehensive and standardised assessment process.

4) Rating accuracy: Considering the inherently quantitative nature of peer assessment, which involves rating peers' work using a rating scale, the study relies on the overarching concept of rating accuracy to evaluate the precision of peer assessment. Rating accuracy pertains to the extent to which a rater offers precise and accurate scores or ratings while evaluating a performance or work. It signifies the proximity or divergence between the established criterion ratings, which serve as the standards, and the actual ratings assigned by the rater, as elucidated by Wang and Engelhard (2019). Assessing rating accuracy is crucial as it serves as a metric for gauging the rater's capacity to consistently and reliably evaluate and score performances. By examining rating accuracy, researchers can ascertain the level of agreement between the rater's assessments and the established standards, thereby determining the reliability and consistency of the peer assessment process. This measure plays a vital role in validating the effectiveness and trustworthiness of the peer assessment method in providing accurate and consistent evaluations of peers' work.

5) Peer feedback roles: In the peer feedback session of this study, participants were assigned one of three different feedback roles: feedback provider, feedback receiver, and feedback outsider. The feedback provider refers to one-third of the participants who take on the role of providing written comments on the writing essays of other students. They are responsible for

offering insights, suggestions, and constructive feedback to help the receivers improve their writing. The feedback receiver represents one-third of the participants who receive written comments for revision on their own writing essays from their peers. They are the recipients of feedback and use the provided comments to revise and enhance the quality of their writing. The feedback outsider, also accounting for one-third of the participants in a peer feedback session, does not actively engage in providing comments on their peers' writing essays nor receive feedback on their own work. They observe the feedback process but do not participate directly, allowing them to gain insights into the feedback exchange without actively contributing as a provider or receiver.

6) Peer feedback orientation: The process of providing and receiving feedback significantly relies on a particular attribute in individuals – openness, referred to as peer feedback orientation (Kasch et al., 2023). This term encapsulates a pivotal aspect of how individuals approach and engage with feedback, influencing both their ability to provide and receive it effectively. The concept of peer feedback orientation encompasses four distinct dimensions: utility, accountability, social awareness, and self-efficacy. These dimensions play a vital role in shaping the overall dynamic of feedback interactions. Importantly, the concept of peer feedback orientation transcends the perspective of just the feedback receiver. It encompasses both the orientation of those providing feedback and those receiving it. This comprehensive view underscores the collaborative and formative nature of the peer feedback process, emphasising its role in supporting individuals' growth and development.

1.11 Chapter Summary

In conclusion, this chapter has provided an overview of the study's context, introduced the problem statement related to peer feedback and peer assessment, and outlined the research objectives, questions and the underlying theoretical framework, encompassing the Development of Evaluative Judgement, Rater Accuracy Models and Peer Feedback Orientation. Furthermore, the chapter has discussed the conceptual framework and clarified key terms, setting the stage for the subsequent chapters of this research. The subsequent chapter will delve into the existing literature, identifying research gaps and paving the way for a deeper examination of the subject matter.

REFERENCES

- Adachi, C., Tai, J. H. M., & Dawson, P. (2018). Academics' perceptions of the benefits and challenges of self and peer assessment in higher education. *Assessment & Evaluation in Higher Education*, 43(2), 294-306.
- Ajjawi, R., Tai, J., Dawson, P., & Boud, D. (2018). Conceptualising evaluative judgement for sustainable assessment in higher education. In D. Boud, R. Ajjawi, P. Dawson, & J. Tai (Eds.), *Developing evaluative judgement in higher education: Assessment for knowing and producing quality work*, (pp. 7–17). London: Routledge.
- Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus*, 15(2), 1-4.
- Allen, D., & Katayama, A. (2016). Relative second language proficiency and the giving and receiving of written peer feedback. *System*, 56, 96-106.
- Allen, D., & Mills, A. (2016). The impact of second language proficiency in dyadic peer feedback. *Language Teaching Research*, 20(4), 498-513.
- Alqassab, M., Strijbos, J. W., & Ufer, S. (2019). Preservice mathematics teachers' beliefs about peer feedback, perceptions of their peer feedback message, and emotions as predictors of peer feedback accuracy and comprehension of the learning task. *Assessment & Evaluation in Higher Education*, 44(1), 139-154.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington: AERA.
- Anders, B. A. (2023). Is using ChatGPT cheating, plagiarism, both, neither, or forward thinking?. *Patterns*, 4 (3), 100694.
- Aryadoust, V. (2012). Evaluating the psychometric quality of an ESL placement test of writing: A many-facets Rasch study. *Linguistics Journal*, 6(1), 8-33.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bakhtiar, A., Webster, E. A., & Hadwin, A. F. (2018). Regulation and socio-emotional interactions in a positive and a negative group climate. *Metacognition and Learning*, 13(1), 57-90.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191-215.

- Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology*, 4(3), 359-373.
- Bandura, A. (1994). Self-efficacy. In: VS Ramachaudran. *Encyclopedia of Human Behavior*, 4(4), 71-81.
- Basturk, R. (2008). Applying the many-facet Rasch model to evaluate PowerPoint presentation performance in higher education. *Assessment & Evaluation in Higher Education*, 33(4), 431-444.
- Bearman, M. (2018). Prefigurement, identities and agency: The disciplinary nature of evaluative judgement. In D. Boud, R. Ajjawi, P. Dawson, & J. Tai (Eds.), *Developing evaluative judgement in higher education: Assessment for knowing and producing quality work*, (pp. 147–155). London: Routledge.
- Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, 1-13.
- Berg, E. C. (1999). The effects of trained peer response on ESL students' revision types and writing quality. *Journal of Second Language Writing*, 8(3), 215-241.
- Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education*, 94(5), 765-793.
- Black, P., & William, D. (2015). Assessment and classroom learning, assessment in education. *Pilet & Horoks-Actes du Séminaire National de l'ARDM-mars*, 5(1), 7-74.
- Boud, D. (1989). The role of self-assessment in student grading. *Assessment in Higher Education*, 14(1), 20-30.
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151–167.
- Boud, D., & Falchikov, N. (2005). Redesigning assessment for learning beyond higher education. *Research and Development in Higher Education*, 28, 34-41.
- Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education*, 38(6), 698–712.
- Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400–413.

- Boud, D., Dawson, P., Bearman, M., Bennett, S., Joughin, G., & Molloy, E. (2018). Reframing assessment research: Through a practice perspective. *Studies in Higher Education*, 43(7), 1107-1118.
- Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. London: Sage Publications.
- Braine, G. (2003). From a teacher-centered to a student-centered approach: A study of peer feedback in Hong Kong writing classes. *Journal of Asian Pacific Communication*, 13(2), 269-288.
- Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago: University of Chicago Press.
- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, 62(3), 193-217.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research*, 65(3), 245-281.
- Cao, Z., Yu, S., & Huang, J. (2019). A qualitative inquiry into undergraduates' learning from giving and receiving peer feedback in L2 writing: Insights from a case study. *Studies in Educational Evaluation*, 63, 102-112.
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315-1325.
- Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education*, 36(4), 395-407.
- Chang, C. Y. H. (2016). Two decades of research in L2 peer review. *Journal of Writing Research*, 8(1), 81-117.
- Cheah, P. K., Diong, F. W., & Yap, Y. O. (2018). Peer assessment in higher education: Using Hofstede's cultural dimensions to identify perspectives of Malaysian Chinese students. *Pertanika Journal of Social Sciences & Humanities*, 26(3), 1471-1489.
- Chen, C. W. Y. (2010). Graduate students' self-reported perspectives regarding peer feedback and feedback from writing consultants. *Asia Pacific Education Review*, 11(2), 151-158.
- Cheng, W., & Warren, M. (1997). Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higher Education*, 22(2), 233-239.
- Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language testing*, 22(1), 93-121.

- Cheng, X., & Zhang, L. J. (2024). Engaging secondary school students with peer feedback in L2 writing classrooms: A mixed-methods study. *Studies in Educational Evaluation*, 81, 101337.
- Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction*, 20(4), 328–338.
- Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, 98(4), 891–901.
- Cho, Y. H., & Cho, K. (2011). Peer reviewers learn from giving comments. *Instructional Science*, 39(5), 629-643.
- Chong, S. W. (2021). University students' perceptions towards using exemplars dialogically to develop evaluative judgement: The case of a high-stakes language test. *Asian-Pacific Journal of Second and Foreign Language Education*, 6(1), 1-27.
- Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, 24(2), 205-249.
- Cognitive perspectives on rater-mediated assessments using accuracy ratings. *Psychological Test and Assessment Modeling*, 60(1), 33–52.
- Connor, U., & Asenavage, K. (1994). Peer response groups in ESL writing classes: How much impact on revision?. *Journal of Second Language Writing*, 3(3), 257-276.
- Cooksey, R. W. (1996). The methodology of social judgement theory. *Thinking and Reasoning* 2(2), 141–174.
- Crabtree, B. F., & Miller, W. F. (1992). A template approach to text analysis: Developing and using codebooks. In B. F. Crabtree & W. L. Miller (Eds.), *Doing qualitative research* (pp. 93–109). London: Sage Publications.
- Creswell, J. W., & Clark, V. L. P. (2011). *Mixed methods research*. London: SAGE Publications.
- Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. London: Sage Publications.
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. London: Sage Publications.
- Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, 39(3), 124-130.

- Dawson, P., Henderson, M., Mahoney, P., Phillips, M., Ryan, T., Boud, D., & Molloy, E. (2019). What makes for effective feedback: Staff and student perspectives. *Assessment & Evaluation in Higher Education*, 44(1), 25-36.
- De Grez, L., Valcke, M., & Roozen, I. (2012). How effective are self-and peer assessment of oral presentation skills compared with teachers' assessments?. *Active Learning in Higher Education*, 13(2), 129-142.
- De Guerrero, M. C., & Villamil, O. S. (2000). Activating the ZPD: Mutual scaffolding in L2 peer revision. *The Modern Language Journal*, 84(1), 51-68.
- Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory into Practice*, 55(2), 153-159.
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331-350.
- Dong, Y., & Shi, L. (2021). Using Grammarly to support students' source-based writing practices. *Assessing Writing*, 50, 100564.
- Eckes, T. (2008). Rater types in writing performance assessments: A classification approach to rater variability. *Language Testing*, 25(2), 155-185.
- Eckes, T. (2009). Many-facet Rasch measurement. In S. Takala (Ed.), *Reference supplement to the manual for relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment* (Section H). Council of Europe/Language Policy Division.
- Eckes, T. (2015). *Introduction to Many-Facet Rasch Measurement: Analyzing and Evaluating Rater-Mediated Assessments*. (2nd ed.). Frankfurt: Peter Lang.
- Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist*, 46(1), 6-25.
- Engelhard, G., Jr. (1992). The measurement of writing ability with a many-faceted Rasch model. *Applied measurement in Education*, 5(3), 171-191.
- Engelhard, G., Jr. (1994). Examining rater errors in the assessment of written composition with a many-faceted Rasch model. *Journal of Educational Measurement*, 31(2), 93-112.
- Engelhard, G., Jr. (1996). Evaluating rater accuracy in performance assessments. *Journal of Educational Measurement*, 33, 56-70.

- Engelhard, G., Jr. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. Oxfordshire: Routledge.
- Engelhard, G., Jr., & Myford, C. M. (2009). Comparison of single-and double-assessor scoring designs for the assessment of accomplished teaching. *Journal of Applied Measurement*, 10(1), 52-69.
- Engelhard, G., Jr., & Wind, S. (2017). *Invariant measurement with raters and rating scales: Rasch models for rater-mediated assessments*. Oxfordshire: Routledge.
- Engelhard, Jr, G., Jr., Wang, J., & Wind, S. A. (2018). A tale of two models: Psychometric and cognitive perspectives on rater-mediated assessments using accuracy ratings. *Psychological Test and Assessment Modeling*, 60(1), 33-52.
- Er, E., Dimitriadis, Y., & Gašević, D. (2021). A collaborative learning approach to dialogic peer feedback: A theoretical framework. *Assessment & Evaluation in Higher Education*, 46(4), 586-600.
- Esfandiari, R., & Myford, C. M. (2013). Severity differences among self-assessors, peer-assessors, and teacher assessors rating EFL essays. *Assessing Writing*, 18(2), 111-131.
- Esterhazy, R., & Damşa, C. (2019). Unpacking the feedback process: An analysis of undergraduate students' interactional meaning-making of feedback comments. *Studies in Higher Education*, 44(2), 260–274.
- Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research*, 83(1), 70-120.
- Evans, C. (2016). *Enhancing assessment feedback practice in higher education: The EAT framework*. Southampton: University of Southampton.
- Falchikov, N. (2013). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. Oxfordshire: Routledge.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3), 287-322.
- Fan, Y., & Xu, J. (2020). Exploring student engagement with peer feedback on L2 writing. *Journal of Second Language Writing*, 50, 1-13.
- Farrokhi, F., Esfandiari, R., & Schaefer, E. (2012). A many-facet Rasch measurement of differential rater severity/leniency in three types of assessment. *Japan Association for Language Teaching Journal*, 34(1), 79-101.

- Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, 5(1), 80-92.
- Filius, R. M., de Kleijn, R. A., Uijl, S. G., Prins, F. J., van Rijen, H. V., & Grobbee, D. E. (2018). Strengthening dialogic peer feedback aiming for deep learning in SPOCs. *Computers & Education*, 125, 86-100.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, 34(10), 906–911.
- Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986). Detection, diagnosis, and the strategies of revision. *College Composition and Communication*, 37(1), 16-55.
- Forsythe, A., & Johnson, S. (2017). Thanks, but no-thanks for the feedback. *Assessment & Evaluation in Higher Education*, 42(6), 850-859.
- Fowler, Y. (2007). Using peer and self-assessment in interpreter training. *The Critical Link 4: Professionalisation of interpreting in the community*, 70(253-262).
- Freeman, M. (1995). Peer assessment by groups of group work. *Assessment & Evaluation in Higher Education*, 20(3), 289-301.
- Fukazawa, M. (2010). Validity of peer assessment of speech performance. *ARELE: Annual Review of English Language Education in Japan*, 21, 181-190.
- Gao, Y., An, Q., & Schunn, C. D. (2023). The bilateral benefits of providing and receiving peer feedback in academic writing across varying L2 proficiency. *Studies in Educational Evaluation*, 77, 101252.
- Gao, Y., Schunn, C. D. D., & Yu, Q. (2019). The alignment of written peer feedback with draft problems and its impact on revision in peer assessment. *Assessment & Evaluation in Higher Education*, 44(2), 294-308.
- Gere, A. R. (1987). *Writing groups: History, theory, and implications*. Carbondale: Southern Illinois University Press.
- Gibbs, G. R. (2007). Thematic coding and categorizing. *Analyzing Qualitative Data*, 703, 38-56.
- Gielen, M., & De Wever, B. (2015). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior*, 52, 315-325.

- Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, 20(4), 304-315.
- Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social Problems*, 12(4), 436-445.
- Goldfinch, J.M., & Raeside, R. (1990). Development of a peer assessment technique for obtaining individual marks on a group project. *Assessment & Evaluation in Higher Education*, 15(3), 210-225.
- Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review*, 62(4), 255-262.
- Han, C. (2017). Using analytic rating scales to assess English/Chinese bi-directional interpretation: A longitudinal Rasch analysis of scale utility and rater behavior. *Linguistica Antverpiensia, New Series—Themes in Translation Studies*, 16, 196-215.
- Han, C. (2018). Latent trait modelling of rater accuracy in formative peer assessment of English-Chinese consecutive interpreting. *Assessment & Evaluation in Higher Education*, 43(6), 979-994.
- Han, C., & Zhao, X. (2021). Accuracy of peer ratings on the quality of spoken-language interpreting. *Assessment & Evaluation in Higher Education*, 46(8), 1299-1313.
- Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing*, 30, 31-44.
- Harris, J. R. (2011). Peer assessment in large undergraduate classes: An evaluation of a procedure for marking laboratory reports and a review of related practices. *Advances in Physiology Education*, 35(2), 178-187.
- Harrison, C. J., Könings, K. D., Schuwirth, L., Wass, V., & Van der Vleuten, C. (2015). Barriers to the uptake and use of feedback in the context of summative assessment. *Advances in Health Sciences Education*, 20(1), 229-245.
- Hartley, A., Mason, I., Peng, G., & Perez, I. (2003). Peer-and self-assessment in conference interpreter training. Pedagogical Research Fund in Languages, Linguistics and Area Studies. Retrieved on June, 23, 2023 from <https://www.llas.ac.uk/projects/1454.html>.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.

- Hasim, Z. (2014). An integration of a process approach and formative assessment into the development of teaching and learning of ESL writing in a Malaysian university: A sociocultural perspective (Doctoral dissertation, University of Waikato).
- Henderson, M., Boud, D., Molloy, E., Dawson, P., Phillips, M., Ryan, T., & Mahoney, P. (2018). *Feedback for learning: Closing the assessment loop*. Department of Education and Training.
- Hepplestone, S., & Chikwa, G. (2016). Exploring the processes used by students to apply feedback. *Student Engagement and Experience Journal*, 5(1), 1-15.
- Hounsell, D. (2007). Towards more sustainable feedback to students. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education: Learning for the longer term* (pp. 111-123). Oxfordshire: Routledge.
- Hu, G., & Lam, S. T. E. (2010). Issues of cultural appropriateness and pedagogical efficacy: Exploring peer review in a second language writing class. *Instructional Science*, 38(4), 371-394.
- Huang, Y., & Zhang, L. J. (2022). Facilitating L2 writers' metacognitive strategy use in argumentative writing using a process-genre approach. *Frontiers in Psychology*, 13, 1036831.
- Huisman, B., Saab, N., Van den Broek, P., & Van Driel, J. (2019). The impact of formative peer feedback on higher education students' academic writing: a Meta-Analysis. *Assessment & Evaluation in Higher Education*, 44(6), 863-880.
- Huisman, B., Saab, N., Van Driel, J., & Van Den Broek, P. (2018). Peer feedback on academic writing: undergraduate students' peer feedback role, peer feedback perceptions and essay performance. *Assessment & Evaluation in Higher Education*, 43(6), 955-968.
- Huisman, B., Saab, N., Van Driel, J., & Van Den Broek, P. (2020). A questionnaire to assess students' beliefs about peer-feedback. *Innovations in Education and Teaching International*, 57(3), 328-338.
- Hulleman, C. S., Durik, A. M., Schweigert, S. A., & Harackiewicz, J. M. (2008). Task values, achievement goals, and interest: An integrative analysis. *Journal of Educational Psychology*, 100(2), 398-416.
- Husin, M. S., & Ariffin, K. (2008). Students' perceptions towards peer feedback approach in a Malaysian English as a second language (ESL) classroom. *Jurnal Gading UiTM Pahang*, 12(2), 19-34.
- Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, 39(2), 83-101.

- Jonsson, A. (2013). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education*, 14(1), 63-76.
- Kasch, J., Van Rosmalen, P., & Kalz, M. (2023). A thematic analysis of factors influencing student's peer-feedback orientation. In O. Noroozi & B. De Wever (Eds.), *The power of peer learning: Social interaction in learning and development* (pp. 265-282). Cham: Springer.
- Kasch, J., Van Rosmalen, P., Henderikx, M., & Kalz, M. (2021). The factor structure of the peer-feedback orientation scale (PFOS): Toward a measure for assessing students' peer-feedback dispositions. *Assessment & Evaluation in Higher Education*, 47(1), 15-28.
- Khan, A., Hassan, N., & Ali, A. (2023). Implementing formative assessment in Malaysia: Teachers' viewpoints. *Language Testing in Focus: An International Journal*, 7, 28-41.
- Kim, M. (2009). The impact of an elaborated assessee's role in peer assessment. *Assessment & Evaluation in Higher Education*, 34(1), 105-114.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284.
- Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*, 44, 100450.
- Kovach, R. A., Resch, D. S., & Verhulst, S. J. (2009). Peer assessment of professionalism: A five-year experience in medical clerkship. *Journal of General Internal Medicine*, 24, 742-746.
- Kwan, K., & Leung, R. (1996). Tutor versus peer group assessment of student performance in a simulation training exercise. *Assessment & Evaluation in Higher Education*, 21(3), 205-215.
- Lam, R. (2010). A peer review training workshop: Coaching students to give and evaluate peer feedback. *TESL Canada Journal*, 27(2), 114.
- Langan, A. M., Shuker, D. M., Cullen, W. R., Penney, D., Preziosi, R. F., & Wheeler, C. P. (2008). Relationships between student characteristics and self-, peer and tutor evaluations of oral presentations. *Assessment & Evaluation in Higher Education*, 33(2), 179-190.
- Larsen-Freeman, D. (2016). Classroom-oriented research from a complex systems perspective. *Studies in Second Language Learning and Teaching*, 6(3), 377-393.

- Latifi, S., Noroozi, O., & Talaee, E. (2020). Worked example or scripting? Fostering students' online argumentative peer feedback, essay writing and learning. *Interactive Learning Environments*, 31(2), 655-669.
- Latifi, S., Noroozi, O., Hatami, J., & Biemans, H. J. A. (2021). How does online peer feedback improve argumentative essay writing and learning? *Innovations in Education and Teaching International*, 58(2), 195–206.
- Leary, M. R., & Baumeister, R. F. (1995). The need to belong. *Psychological Bulletin*, 117(3), 497-529.
- Lee, S. B. (2017). University students' experience of 'scale-referenced' peer assessment for a consecutive interpreting examination. *Assessment & Evaluation in Higher Education*, 42(7), 1015-1029.
- Leki, I. (1990). Potential problems with peer responding in ESL writing classes. *CATESOL Journal*, 3(1), 5-19.
- Li, H., Xiong, Y., Zang, X., L. Kornhaber, M., Lyu, Y., Chung, K. S., & K. Suen, H. (2016). Peer assessment in the digital age: A meta-analysis comparing peer and teacher ratings. *Assessment & Evaluation in Higher Education*, 41(2), 245-264.
- Li, J., & De Luca, R. (2014). Review of assessment feedback. *Studies in Higher Education*, 39(2), 378–393.
- Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology*, 41(3), 525-536.
- Lim, H. O. (2013). A jury of your peers: Peer assessment in consecutive interpretation pedagogy. *International Journal of Interpretation and Translation*, 11(1), 77-97.
- Linacre, J. M. (1989/1994). *Many-facet Rasch measurement*. San Diego: MESA Press.
- Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean. *Rasch Measurement Transactions*, 16(2), 878.
- Linacre, J. M. (2013). A user's guide to FACETS: Program manual 3.71.2. Retrieved on June, 23, 2023 from <http://www.winsteps.com/a/facets-manual.pdf>.
- Linderbaum, B. A., & Levy, P. E. (2010). The development and validation of the Feedback Orientation Scale (FOS). *Journal of Management*, 36(6), 1372–1405.
- Linnenbrink, E. A., & Pintrich, P. R. (2003). The role of self-efficacy beliefs instudent engagement and learning in the classroom. *Reading & Writing Quarterly*, 19(2), 119-137.

- Liu, C., & Yu, S. (2022). Reconceptualizing the impact of feedback in second language writing: A multidimensional perspective. *Assessing Writing*, 53, 100630.
- Liu, J., & Edwards, J. G. H. (2018). *Peer response in second language writing classrooms*. Ann Arbor: University of Michigan Press.
- Liu, N. F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education*, 11(3), 279-290.
- Lizzio, A., & Wilson, K. (2008). Feedback on assessment: Students' perceptions of quality and effectiveness. *Assessment & Evaluation in Higher Education*, 33(3), 263-275.
- London, M., & Smither, J. W. (2002). Feedback orientation, feedback culture, and the longitudinal performance management process. *Human Resource Management Review*, 12(1), 81-100.
- Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of second language writing*, 18(1), 30-43.
- Lunz, M. E. (1992). New ways of thinking about reliability. *Professions Education Researcher Quarterly*, 13(4), 16-18.
- Luo, J., & Chan, C. K. Y. (2023a). Exploring the process of evaluative judgement: The case of engineering students judging intercultural competence. *Assessment & Evaluation in Higher Education*, 48(7), 951-965.
- Luo, J., & Chan, C. K. Y. (2023b). Conceptualising evaluative judgement in the context of holistic competency development: Results of a Delphi study. *Assessment & Evaluation in Higher Education*, 48(4), 513-528.
- Lynch, D. H., & Golen, S. (1992). Peer evaluation of writing in business communication classes. *Journal of Education for Business*, 68(1), 44-48.
- Magin, D., & Helmore, P. (2001). Peer and teacher assessments of oral presentation skills: How reliable are they?. *Studies in Higher Education*, 26(3), 287-298.
- Malaysian Examinations Council. (2019). *Malaysian University English Test (MUET): Regulations and test specifications*. Selangor: Malaysian Examinations Council.
- Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, 26(1), 75-100.

- McConlogue, T. (2012). But is it fair? Developing students' understanding of grading complex written work through peer assessment. *Assessment & Evaluation in Higher Education*, 37(1), 113–123.
- McConlogue, T. (2015). Making judgements: Investigating the process of composing and receiving peer feedback. *Studies in Higher Education*, 40(9), 1495–1506.
- McNamara, T., Knoch, U., & Fan, J. (2019). *Fairness, justice, and language assessment: The role of measurement*. Oxford: Oxford University Press.
- Merry, S., Price, M., Carless, D., & Taras, M. (2013). *Reconceptualising feedback in higher education*. Oxfordshire: Routledge.
- Min, H. T. (2005). Training students to become successful peer reviewers. *System*, 33(2), 293–308.
- Min, H. T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing*, 15(2), 118–141.
- Moore, C., & Teather, S. (2013). Engaging students in peer review: Feedback as learning. *Issues in Educational Research*, 23(2), 196–211.
- Moore, L., Graham, A., & Diamond, I. (2003). On the feasibility of conducting randomised trials in education: Case study of a sex education intervention. *British Educational Research Journal*, 29(5), 673–689.
- Mowl, G., & Pain, R. (1995). Using self and peer assessment to improve students' essay writing: A case study from geography. *Innovations in Education and Training International*, 32(4), 324–335.
- Mulliner, E., & Tucker, M. (2017). Feedback on feedback practice: Perceptions of students and academics. *Assessment & Evaluation in Higher Education*, 42(2), 266–288.
- Nelson, G. L., & Murphy, J. M. (1993). Peer response groups: Do L2 writers use peer comments in revising their drafts?. *TESOL Quarterly*, 27(1), 135–141.
- Nelson, R. (2018). Barriers to the cultivation of evaluative judgement: A critical and historical perspective. In D. Boud, R. Ajjawi, P. Dawson, & J. Tai (Eds.), *Developing evaluative judgement in higher education: Assessment for knowing and producing quality work*, (pp. 51–59). London: Routledge.
- Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, 35(5), 501–517.

- Nicol, D. (2013). Resituating feedback from the reactive to the proactive. In D. Boud & E. Molloy (Eds.), *Feedback in higher and professional education: Understanding and doing it well* (pp. 34–49). Oxfordshire: Routledge.
- Nicol, D. (2014). Guiding principles for peer review: Unlocking learners' evaluative skills. In C. Kreber, C. Anderson, N. Entwistle, & J. McArthur (Eds.), *Advances and innovations in university assessment and feedback* (pp. 197–224). Edinburgh: Edinburgh University Press.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218.
- Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education*, 39(1), 102–122.
- O'Neill, T. R., & Lunz, M. E. (1996, April). Examining the invariance of rater and project calibrations using a multi-facet Rasch model. Paper presented at the Annual Meeting of the American Educational Research Association, New York.
- Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education*, 25(1), 23-38.
- Pajares, F. (2002). Gender and perceived self-efficacy in self-regulated learning. *Theory into Practice*, 41(2), 116-125.
- Palmer, B., & Major, C. H. (2008). Using reciprocal peer review to help graduate students develop scholarly writing skills. *The Journal of Faculty Development*, 22(3), 163.
- Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 247–266). Oxfordshire: Routledge.
- Patchan, M. M., & Schunn, C. D. (2015). Understanding the benefits of providing peer feedback: How students respond to peers' texts of varying quality. *Instructional Science*, 43, 591-614.
- Percival, F., & Ellington, H. (1988). *A handbook of educational technology*. (2nd ed.). London: Kogan Page.
- Pope, N. K. L. (2005). The impact of stress in self- and peer assessment. *Studies in Higher Education*, 30(1), 51-63.

- Postman, L., & Tolman, E. C. (1959). Brunswik's probabilistic functionalism. In S. Koch (Ed.), *Psychology: The study of a science, Vol. 1: Sensory, perceptual, and physiological formulations*. New York: McGraw-Hill.
- Price, M., Handley, K., & Millar, J. (2011). Feedback: Focusing attention on engagement. *Studies in Higher Education*, 36(8), 879-896.
- Quinton, S., & Smallbone, T. (2010). Feeding forward: Using feedback to promote student reflection and learning—a teaching model. *Innovations in Education and Teaching International*, 47(1), 125-135.
- Rahimi, M., & Zhang, L. J. (2022). Effects of an engaging process-genre approach on student engagement and writing achievements. *Reading & Writing Quarterly*, 38(5), 487-503.
- Reeve, J., & Tseng, C. M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, 36(4), 257-267.
- Rice, P. L., & Ezzy, D. Qualitative Research Methods: A Health Focus. *Chronic Diseases in Canada*, 21(3), 137.
- Ryan, T., & Henderson, M. (2018). Feeling feedback: Students' emotional responses to educator feedback. *Assessment & Evaluation in Higher Education*, 43(6), 880-892.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-144.
- Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35(5), 535-550.
- Sadler, D. R. (2013). Assuring academic achievement standards: From moderation to calibration. *Assessment in Education: Principles, Policy & Practice*, 20(1), 5-19.
- Saito, H. (1994). Teachers' practices and students' preferences for feedback on second language writing: A case study of adult ESL learners. *TESL Canada Journal*, 11(2), 46-70.
- Saito, H., & Fujita, T. (2004). Characteristics and user acceptance of peer rating in EFL writing classrooms. *Language Teaching Research*, 8(1), 31-54.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189.
- Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology*, 8, 1-11.

- Speyer, R., Pilz, W., Van Der Kruis, J., & Brunings Wouter, J. (2011). Reliability and validity of student peer assessment in medical education: A systematic review. *Medical Teacher*, 33(11), 572–585.
- Srichanyachon, N. (2012). An investigation of university EFL students' attitudes toward peer and teacher feedback. *Educational Research and Reviews*, 7(26), 558-562.
- Strijbos, J. W., Narciss, S., & Dünnebie, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency?. *Learning and Instruction*, 20(4), 291-303.
- Sullivan, K., & Hall, C. (1997). Introducing students to self-assessment. *Assessment & Evaluation in Higher Education*, 22(3), 289-306.
- Sulsky, L. M., & Balzer, W. K. (1988). Meaning and measurement of performance rating accuracy: Some methodological and theoretical concerns. *Journal of Applied Psychology*, 73(3), 497.
- Suñol, J. J., Arbat, G., Pujol, J., Feliu, L., Fraguell, R. M., & Planas-Lladó, A. (2016). Peer and self-assessment applied to oral presentations from a multidisciplinary perspective. *Assessment & Evaluation in Higher Education*, 41(4), 622-637.
- Sutton, P., & Gill, W. (2010). Engaging Feedback: Meaning, Identity and Power. *Practitioner Research in Higher Education*, 4(1), 3-13.
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher education*, 76, 467-481.
- Tai, J., Canny, B. J., Haines, T. P., & Molloy, E. K. (2016). The role of peer-assisted learning in building evaluative judgement: Opportunities in clinical medical education. *Advances in Health Sciences Education*, 21, 659-676.
- The University of Reading. (2009). TEEP extended textbook. Berkshire: The University of Reading. Retrieved on June, 23, 2023 from http://www.cals.rdg.ac.uk/teep/files/2001_teep_extended_handbook.pdf
- To, J., Panadero, E., & Carless, D. (2022). A systematic review of the educational uses and effects of exemplars. *Assessment & Evaluation in Higher Education*, 47(8), 1167-1182.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249-276.
- Topping, K. J. (2009). Peer assessment. *Theory into Practice*, 48(1), 20-27.

- Tsui, A. B., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments?. *Journal of Second Language Writing*, 9(2), 147-170.
- Van den Berg, I., Admiraal, W., & Pilot, A. (2006). Design principles and outcomes of peer assessment in higher education. *Studies in Higher education*, 31(03), 341-356.
- Van Popta, E., Kral, M., Camp, G., Martens, R. L., & Simons, P. R. J. (2017). Exploring the value of peer feedback in online learning for the provider. *Educational Research Review*, 20, 24-34.
- Vasu, K., Ling, C. H., & Nimehchisalem, V. (2016). Malaysian tertiary level ESL students' perceptions toward teacher feedback, peer feedback and self-assessment in their writing. *International Journal of Applied Linguistics and English Literature*, 5(5), 158-170.
- Villamil, O. S., & De Guerrero, M. C. (1996). Peer revision in the L2 classroom: Social-cognitive activities, mediating strategies, and aspects of social behavior. *Journal of Second Language Writing*, 5(1), 51-75.
- Vrugt, A., & Oort, F. J. (2008). Metacognition, achievement goals, study strategies and academic achievement: Pathways to achievement. *Metacognition and Learning*, 3, 123-146.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.
- Wang, J., & Engelhard Jr, G. (2019). Conceptualizing rater judgments and rating processes for rater-mediated assessments. *Journal of Educational Measurement*, 56(3), 582-609.
- Wang, J., Engelhard Jr, G., & Wolfe, E. W. (2016). Evaluating rater accuracy in rater-mediated assessments using an unfolding model. *Educational and Psychological Measurement*, 76(6), 1005-1025.
- Wang, J., Engelhard Jr, G., Raczynski, K., Song, T., & Wolfe, E. W. (2017). Evaluating rater accuracy and perception for integrated writing assessments using a mixed-methods approach. *Assessing Writing*, 33, 36-47.
- Weigle, S. C. (1994). Effects of training on raters of ESL compositions. *Language Testing*, 11(2), 197-223.
- Weir, C. (1990). *Communicative language testing*. London: Prentice-Hall.
- Wind, S. A. (2014). Examining rating scales using Rasch and Mokken models for rater-mediated assessments. *Journal of Applied Measurement*, 15(2), 100-132.

- Wind, S. A., & Engelhard Jr, G. (2012). Examining rating quality in writing assessment: Rater agreement, error, and accuracy. *Journal of applied measurement*, 13(4), 321-335.
- Wind, S. A., & Engelhard Jr, G. (2013). How invariant and accurate are domain ratings in writing assessment?. *Assessing Writing*, 18(4), 278-299.
- Wind, S. A., Wolfe, E. W., Engelhard Jr, G., Foltz, P., & Rosenstein, M. (2018). The influence of rater effects in training sets on the psychometric quality of automated scoring for writing assessments. *International Journal of Testing*, 18(1), 27-49.
- Winstone, N. E., & Nash, R. A. (2016). *The developing engagement with feedback toolkit (DEFT)*. York: Higher Education Academy.
- Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, 52(1), 17-37.
- Wolfe, E. W. (2014). *Methods for monitoring rating quality: Current practices and suggested changes*. London: Pearson.
- Wolfe, E. W., & McVay, A. (2012). Application of latent trait models to identifying substantively interesting raters. *Educational Measurement: Issues and Practice*, 31(3), 31-37.
- Wolfe, E. W., Jiao, H., & Song, T. (2015). A Family of Rater Accuracy Models. *Journal of Applied Measurement*, 16(2), 153-160.
- Yu, S., & Lee, I. (2016). Peer feedback in second language writing (2005-2014). *Language Teaching*, 49(4), 461-493.
- Zhu, Q., & Carless, D. (2018). Dialogue within peer feedback processes: Clarification and negotiation of meaning. *Higher Education Research & Development*, 37(4), 883-897.
- Zhu, Q., & To, J. (2021). Proactive receiver roles in peer feedback dialogue: Facilitating receivers' self-regulation and co-regulating providers' learning. *Assessment & Evaluation in Higher Education*, 47(8), 1200-1212.
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation: Theory, research and applications* (pp. 13-39). San Diego: Academic Press.
- Zimmerman, B. J., & Kitsantas, A. (1999). Acquiring writing revision skill: Shifting from process to outcome self-regulatory goals. *Journal of Educational Psychology*, 91(2), 241.