March 2023

L2 Writer Identity Construction in Academic Written Discourse: A Multi-case Study

Beibei Ren
University of South Florida

L2 Writer Identity Construction in Academic Written Discourse: A Multi-case Study

by

Beibei Ren

A dissertation submitted in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
Department of World Languages
College of Arts and Sciences
University of South Florida

Major Professor: Wei Zhu, Ph.D.
Brandon Tullock, Ph.D.
Nicole Tracy-Ventura, Ph.D.
Matt Kessler, Ph.D.

Date of Approval:
March 8, 2023

Keywords: Writer Identity, L2 Doctoral Students, Social Constructivism, Context, Agency

DEDICATION

This dissertation is dedicated to my parents, whose love has always been a source of courage, light, and power in my life.

ACKNOWLEDGMENTS

Looking back on the five years since I came to the U.S. to pursue my doctoral degree, I cannot express with words what a bittersweet journey it has been. This path was not without its frustrations, homesickness, and loneliness, but all these hardships have led me to become a stronger person. I appreciate all the experiences that have molded me to be who I am today. From being a student who was completely at a loss after arriving at a foreign country to being the author of this dissertation, I have seen that there are so many people that have played an indispensable role in this process. Without them, I would not have been able to make it through.

First and foremost, I would like to extend my deepest and most sincere gratitude to my advisor, Dr. Wei Zhu, without whom the whole doctoral journey would have been impossible. In the past five years, she has walked me through the whole process at my own pace with great patience and understanding. I am incredibly fortunate to have her as my PhD advisor. Besides the emotional support, I am also deeply indebted to her for her professional guidance in conducting research. Her conscientiousness and rigor in doing research has greatly shaped my attitude, and that will continuously influence my academia career in the years to come.

I am thankful for my committee members, Dr. Brandon Tullock, Dr. Matt Kessler, and Dr. Nicole Tracy-Ventura, who offered invaluable and insightful feedback on my dissertation proposal and the dissertation itself.

I also want to specially thank Dr. Camilla Vasquez, who encouraged me profoundly in the first year of my PhD. It meant so much to me and motivated me to proceed with my studies.

Thanks also go to my participants, who have volunteered their time to participate in this study and to generously share with me their papers and their thoughts on the papers, providing invaluable data for this dissertation research.

I also appreciate Dr. Eric Shepherd and Dr. Xizhen Qin, two mentors from the Chinese program, who have greatly helped me grow as a Chinese instructor. I thank Dr. Judith Bridges for her help and support with my teaching of the linguistics course.

I am grateful that I was awarded the Dissertation Completion Fellowship in Fall 2022. The grant helped me tremendously during the completion of this dissertation.

I am also grateful for my colleagues and friends from the LALS program who have been great companions in this PhD journey. It is them who made my monotonous academic life fun and hopeful. Oksana Bomba, Anna Stepanyan—I will always remember the beautiful sunrise and sunsets we have seen together in Florida. Özge Güney, thank you for being great company, and thank you for always thinking of me. Joseph Rasico, thank you for your support and encouragement in the first year of my PhD. I also want to thank Dr. Jining Han for the generous help he offered me throughout this whole journey. Thanks also go to Dr. Yanjia Zhang for being a wonderful friend of mine.

And, my cat, Mianmian—I am so happy to have you in my life. Thank you for your fluffy company 24/7 in the past five years. I love you.

Last but not least, my wholehearted appreciation goes to my parents, Mingqi Ren and Xiaofeng Zhou, and my sister, Mengqiao Ren, for their unconditional love and support in my life. No matter what my choices are, they are always the ones who will be by my side and support me. They are the reasons I have been able to come this far.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Writer identity has gradually become a focus in writing scholarship in recent years. From a social constructivist lens, writer identity is not optional; it resides in all texts. And it is a construct that does not exist in a vacuum, but is shaped by the sociocultural and academic context, and simultaneously individual writers agentively select from the socially available repertoire to construct their identities in text. Departing from social constructivism, this study adopted Ivanic's (1998) conceptualization of writer identity, which consists of *autobiographical self, discoursal self, self as author* and *possibilities for selfhood*, and highlighted the role of agency in the construction of writer identity.

This dissertation employed a multi-case study design, and looked into how three L2 doctoral students studying in the U.S. setting constructed their identities in their academic papers of different kinds and the patterns of identity construction across their papers. Among the three participants, two came from the hard discipline background whereas the other one from soft discipline. Further, this study examined the variables, including contextual factors (*autobiographical self* and *possibilities for selfhood*) and individual agency, that contributed to the three L2 doctoral students' identity construction in their papers. Data collected for this study included three writing samples from each of the students, and three to four rounds of interviews with the participants. Interviews encompassed background interviews, and interviews about the papers. Supplementary data were also collected to provide additional insights into the students' identity construction that included research log, advisors'/reviewers'/co-authors' feedback on the papers, course syllabi and so on.

This dissertation analyzed writer identity through combining Hyland's (2018) metadiscourse model and interactional model (2005). The findings revealed that the three strands of metadiscourse followed the same pattern in all the three students' papers, showing that these three students probably attached more importance to constructing identities as writers using explicit linguistic devices to glue texts together and to signal the internal relations of their text. Comparatively, they tended to project identities less prominently as writers who engaged readers directly in the texts. However, more metadiscourse markers were found in the doctoral student' papers from the soft discipline. This demonstrated the more discursively elaborate nature of the soft discipline, where writers may need more metadiscoursal items to construct their identities as disciplinary insiders. The findings further showed that the three participants' construction of identities through metadiscourse resources was shaped by contextual factors such as field conventions and senior researchers' guidance. Agency, however, played a relatively minor role in the construction of their identities in the academic papers. Insights generated from the study and directions for future research are discussed.

CHAPTER 1: INTRODUCTION

1.1 Statement of the problem

From a social constructivist perspective, identity is a fluid concept (Ivanič, 1998; Matsuda, 2015), and it varies across situational context and takes on idiosyncratic characteristics. Identity is not optional (Flowerdew & Wang, 2015) and it is inscribed and embedded in social actions. The self that is manifested in discourse is a product of both individuals' affiliation to particular beliefs and practices socially available to them and individual s' agency in shaping form and meaning in their texts (Matsuda, 2015). Text represents social reality, and social relations and social identities (Fairclough, 1992). Writer's identities are socially constructed which undergo constant changes and contestations amidst the confluence of the social and academic contexts (Gee, 2000); hence a writer may construct a different identity from one text to another. Identities are multilayered (Ivanič, 1998); in academic writing, identities constructed in texts include "the 'self' that a person brings to the act of writing, the "self" she constructs through the act of writing, and the way in which the writer is perceived by the reader(s) of the writing" (Burgess & Ivanič, 2010, p. 232).

Perceptions of identity through the lens of social constructivism view that individual writers appropriate social conventions available to them to construct identities (Matsuda, 2015). In other words, immediate situational context and the sociocultural context dynamically interact with agency in the construction of identity (Ivanič, 1998; Matsuda, 2015). In writing a piece of text, writers' identities are projected by following rules and conventions from within their discourse communities, yet they simultaneously exercise agency in the choice of linguistic resources (Hyland, 2012a). For instance, a writer can determine the degree of demonstrating the conspicuous presentation of themselves by using self-referenced pronouns or by avoiding that through passive voice. Flowerdew and Wang (2015) pointed out that “[t]o what extent individuals are free to construct their own identity, their agency, and to what extent their identity is controlled by contextual forces, social structure, is an important focus for debate in the literature on identity” (p. 82).

Text production and identity construction are inextricably related to the various institutional, academic, and sociocultural conditions where individual writers are situated (Fairclough, 1992). Hyland (1999) notes that writers’ “discoursal decisions are influenced by, and deeply embedded in, the epistemological and social conventions of their disciplines” (p. 341). In the past couple of decades, the nature of academic writing has been viewed as a social and cultural practice (Atkinson, 2003; Canagarajah, 2000) involving writer’s interactions with text and people from their community of practice (e.g., Fairclough, 1992; Wenger, 1999). Writing includes three interconnected dimensions: the text itself, the social interaction surrounding the production of the writing, and the sociocultural context in which the social interaction is embedded (e.g., Fairclough, 1989; Ivanič & Weldon, 2014). The act of writing contains dialogues in a stream of activities, between or among writers, readers, disciplinary communities and the broader sociocultural context (Wang & Parr, 2021). Academic writing hence is writers’ engagement in the social practices which reflect the prevailing knowledge-making and knowledge-telling conventions constructed by discourse community members (Hyland, 2014). As Tardy (2016) stated, “adopting or performing a discourse is essentially putting on an identity" (p. 352). Through the employment of linguistic markers which could demonstrate who the writers are, writers' identities are emergent in the texts.

However, projecting a credible and scholarly image is challenging, as Kamler and Thompson (2006) claimed, "The text is an extension of the scholar, a putting of 'self' out there which is either successful – or not" (p. 15). Second language (L2) doctoral students, who write academic papers in a language which is not their first language, during the process of being apprenticed into their communities of practice, are expected to carry out research and academic writing under the guidance of their advisor and are "expected to be, or to quickly become, proficient and prolific writers" (Aitchison et al, 2012, p. 435). Research has shown the struggles L2 doctoral students are faced when they are in the process of mastering discipline-specific academic conventions in order to participate in their community of practice as a disciplinary insider (e.g., Sung, 2022; Xu, 2017; Xu & Zhang, 2019); Yet L2 doctoral students may carry unique cultural and linguistic capital (e.g., Bourdieu, 1986) that can make their identity construction characterized by features different from other groups of writers, such as native writers of English. Participating in disciplinary dialogues places a challenge on L2 doctoral students because they are expected to voice themselves appropriately and construct a scholarly identity through writing. Green (2005) argues that doctoral enculturation is "as much about identity formation as it is about knowledge production" (p. 153).

Identity construction is key for doctoral writers to achieve credibility as disciplinary insiders. Claiming competent and professional identities through writing requires doctoral students not only to have a comprehensive grasp of field knowledge but also to be able to critically amalgamate the knowledge, make knowledge contribution, and maintain an authorial stance. Linguistic resources L2 doctoral writers appropriate could manifest their identities constructed through the texts. The current scholarship has investigated writer identity constructed through multiple linguistic features in academic written discourse among different groups of writers (e.g., Burke, 2010; Hyland, 2005, 2010; Ivanič, 1998; Ivanič & Camps, 2001; Ouellette, 2008). Textual analysis on writer identity has generated rich understandings of this construct, yet combining interviews with textual analysis would present a more well-rounded picture of how identities are constructed by the writers. Given this, this dissertation first analyzed identities constructed through the use of linguistic features and conducted interviews with the L2 doctoral participants to elicit emic perspectives regarding their perceptions of the construction of identities through the deployment of certain linguistic markers.

Research writing is a social practice critical to knowledge making and to claiming disciplinary membership and identity (Green, 2005; Hyland, 2004a; Kamler & Thompson, 2006). The expectations placed on doctoral writers are not only completing their dissertations in order to graduate; writing for different genres is also an essential need for doctoral students. For example, writing for publication has increasingly become a common academic activity for doctoral students to enhance personal and institutional performance (Kamler, 2008), and by publishing their research, they can establish their academic status and increase the possibilities of securing jobs after graduation. Doctoral students may also need to complete writing tasks for various other contexts, such as course assignments, scholarship and grant applications (Paré et al., 2011). It is important for doctoral students to compose academic writing of various genres and construct corresponding identities in those texts. Appropriating linguistic resources accepted and preferred by the current discourses can help writers establish intended relationships with their interactants and project desired identities (Hyland, 2010). A few studies have examined L2 doctoral students' self-representation during the completion of their PhD degree (e.g., Botelho de

Magalhães et al., 2019; Hirvela & Belcher, 2001), and a body of literature has examined doctoral writers' identity construction in various texts (e.g., Chiu, 2016; Deng, 2012; Hyland, 2018; Thompson, 2012; Wu & Buripakdi, 2021), which enriches the current body of writer identity scholarship. However, rarely have previous studies given attention to how one individual's writer identities are constructed in writing of different types (i.e., writing of various genres). It is apparent that identity is prone to variations across different situational contexts; investigating writer identity in texts written for different purposes through their linguistic choices can provide in-depth understanding of this construct.

Therefore, how and why L2 doctoral writers select and assemble linguistic resources so that they can contribute to the outcome of writing – why they present themselves this way in this context – is one of the central questions in this dissertation. Specifically, this dissertation explores three L2 doctoral student writers' identity construction in the academic papers they compose during their PhD in an ESL context, and factors that may contribute to their identity construction. Data collected to address the issues encompass three papers of different types from each of the L2 doctoral students, and the additional materials such as research log, comments and feedback on their papers. Three to four rounds of interviews were conducted with each of the participants to seek information about their background and their perceptions, understandings and reasons for employing certain linguistic resources in the papers the way they did. Meanwhile, the three PhD students come from different disciplinary background and stages of studies; in doing so, this dissertation aims to uncover the various contextual factors and their agency contributing to the construction of writer identity construction in the L2 doctoral students' writing of different genres.

1.2 Purpose of the study

Student writers are expected to take on an academic and authoritative voice in the writing of their papers (Swales & Feak, 2004). Previous work has investigated identity constructed in academic written discourse in various settings by concentrating on linguistic features of texts at different levels (e.g., Li & Deng, 2019; Ivanič, 1994, 1998; Ivanič & Camps, 2001; Matsuda, 2001; Ouellette, 2008). L2 doctoral students' identity construction through their papers has not received much attention with regard to discoursal identity constructed in their written texts. Since the demonstration of writer identity is an important index associated with the quality of L2 writing (e.g., Zhao, 2017), for L2 writers, it is significant to understand how self-representation is conveyed through doctoral students' writing and explore their perceptions of this construct, in order to raise their critical language awareness and promote writing proficiency. Additionally, academic writing, as a situated social practice, usually responds to and is shaped by specific communicative context. After all, "it is through our use of community discourses that we claim or resist membership of social groups to define who we are in relation to others" (Hyland, 2010, p. 160). The existent body of research seems to have not systematically explored doctoral students' writer identity in an ESL context. Merriam (1998) propounded that "what is being studied in education is assumed to be in flux, multifaceted, and highly contextual" (p. 206). Therefore, more attention should be directed to the contextual factors that play a role in shaping writer identity in ESL settings, and investigate the dynamic and interactive relationship between writer identity and the factors that shape and constrain its construction.

Framing writer identity in the social constructivist perspective, which views writer identity as being constructed and shaped by externally contextual factors as well as internal factors such as agency, this study aims to explore how identity is constructed from L2 doctoral students' writings. It seeks to uncover textual features that are used to construct these students' identities from texts and the patterns emergent from the writers' papers as well as the patterns across disciplines. It is also concerned with factors, from the perspectives of individual agency, discourse, and genres that contribute to the manifestation of one's identities through written discourse. By investigating writer identity in L2 doctoral students' papers of different kinds, this study hopes to provide insight into writer identity scholarship as well as contribute to the socialization process among L2 doctoral writers into their target discourse community/ies.

1.3 Research questions

This study attempts to explore how L2 doctoral students construct their identities through academic writing in ESL settings. While identity has been a popular research focus in recent years, how L2 doctoral students' identities are constructed in their academic writing of different genres generally have not received much attention. This study, situated in the U.S. context, aims to uncover how writer identity is constructed in L2 doctoral students' writing for different purposes, and how their identity is shaped by contextual and individual factors. More specifically, the research questions are as follows:

1. How do L2 doctoral students construct writer identity in their papers of different genres?

    1a: What metadiscoursal features does their writing have that contribute to the construction of writer identity?

    1b: What are the patterns of metadiscourse use in identity construction in the doctoral students' papers across different written genres?

2. How do contextual factors and agency shape the L2 doctoral students' construction of identity in their papers?

Matsuda suggests (2015, p. 145) “identity is part of the interpersonal meaning that is negotiated through the interaction among the writer and the reader mediated by the text”. Text is the essential component in order for identity to be realized. To understand what kinds of identity are constructed from written discourse, it will be indispensable to examine the discoursal features of a text. Given this, the first research question seeks to examine the linguistic resources L2 doctoral writers utilize when they are engaged with identity construction practices in writing, and the patterns of language use in identity construction across the participants’ papers of different kinds.

From a social constructivist approach, identity is socially constructed and individuals can agentively select their preferred ways of self-representation within the socially acceptable boundaries. Identity construction is subjected to various contextual factors. Considering its fluid and context-dependent nature, the second research question aims to identify factors, both contextual and individual, that influence the construction of identity under particular sociocultural and academic contexts.

1.4 Theoretical framework

This dissertation adopts the social constructivist view of writer identity. Social constructivism accentuates the importance of sociocultural contexts that writers bring with them prior to writing and the immediate learning and writing contexts (Kim, 2001). The former can include academic conventions specific to a discourse community, disciplinary knowledge, the writer’s previous writing and learning experience, and so on, and the latter may entail the specific group of target audience to be addressed, interactions among peers, the requirements for writing, and the goals for writing (i.e., for personal or academic purposes). With the multilayered contexts shaping writer's identity taken into consideration, social constructivism highlights the socially constructed nature of identity.

On the other hand, the social constructivist approach highlights the role of individual agency in shaping the identity in written discourse (e.g., Matsuda, 2015). Matsuda (2001) defines writer identity as "the amalgamative effect of the use of discursive and non-discursive features that language users choose, deliberately or otherwise, from socially available yet ever-changing repertoire" (p. 40). It is clear from this conceptualization of writer identity that though sociocultural variables impact on writer's identity formation, individuals can exercise their agency to select particular linguistic signs in order to achieve the intended effect in writing while rejecting others responding to certain rhetorical circumstances.

Embedded in social constructivist perspectives, this study adopts Ivanič's (1998) conceptual framework of writer identity as the guiding framework. Ivanič (1998) views identity as entailing "aspects of the identity of an actual writer writing a particular text" or "abstract, prototypical identities available in the socio-cultural context of writing" (p. 23). The former is further broken down to "*autobiographical self*", "*discoursal self*", "*self as author/authorial self*", while the latter refers to "*possibilities for self-hood*". The four strands are inextricably related to each other. *Autobiographical self* refers to the life experiences of a writer such as their demographics, education background, academic experience and research interest and so on. *Discoursal self* and *self as author/authorial self* are components of writer identity that are emergent from texts. For instance, the degree of authoritativeness, modesty, and assertiveness can be reflected from writers' linguistic choices, which construct the writers' identities that ultimately could be perceived by readers. *Possibilities for self-hood* concerns the extensive (yet not infinite) pool of identities socially and academically available from which writers can choose, either consciously or subconsciously, to construct their identities in texts. The detailed explanation of the four categories can be found in chapter two.

Though it is noted that writer identity may be difficult to capture in written texts, previous studies have approached writer identity via different sets of linguistic features. This dissertation analyzes L2 doctoral student writers' identity through metadiscourse framework (Hyland, 2018). Metadiscourse, according to Hyland (2018), is "the cover term for the self-reflective expressions used to negotiate interactional meanings in a text, assisting the writer (or speaker) to express a viewpoint and engage with readers as members of a particular community" (p. 44). Metadiscourse further consists of two categories: interactive and interactional metadiscourse, with the latter encompassing stance and engagement markers. Metadiscourse concerns explicit textual devices that signal the relationships between parts of the text and/or between the author and the text and/or between the author and the reader; and the employment of metadiscourse in the unfolding of text is sensitive to context. In the literature review chapter, metadiscourse framework will further be presented.

1.5 Significance of the study

By investigating L2 students' writer identity construction in the ESL setting, this dissertation is significant in the following aspects: First, theoretically, this dissertation adopts Ivanič's (1998) conceptual framework of writer identity, which was initially built upon mature writers' writing who were native speakers of English in the UK setting, and includes four sub-categories of writer identity. This study views writer identity from a social constructivist perspective and claims that identity construction is informed by situation context, and agency simultaneously plays an essential role in the construction of writer identity. This dissertation combines Ivanič's (1998) conceptualization of writer identity and Hyland's (2005, 2018) framework of metadiscourse and interactional model, and examines L2 doctoral students' identity construction in academic papers through metadiscoursal features. In so doing, this research broadens the scope of the applicability of the conceptual framework and extends its theoretical value.

Second, this study looks into writer identity constructed by L2 doctoral writers from different disciplines, and discusses how the employment of metadiscoursal markers demonstrate their identities similarly and differently across disciplines. Further, this dissertation investigates how writer identity might vary across writing of different types written by one L2 doctoral student. The findings could add understandings to the current body of identity research, and shed light on the contextual variables such as participants' disciplinary background and written genres that are associated with writer identity construction.

Third, the study examines the L2 doctoral writer's emic perspectives on identity construction in connection with the textual analysis of writer identity in the doctoral students' academic papers of different genres, and unravels some contextual and personal factors that are indispensable in shaping the construction of the doctoral students' writer identities. For some ESL writers, though they have engaged in academic writing for some time, they may not realize they actually write themselves into the writing (Ivanič, 1998). Hence they largely do not have the critical awareness that the linguistic choices they make would contribute to how readers perceive them as writers, which could highly impact the quality of their writing (Zhao, 2017) and the credibility of their identity being academic writers. Through investigating L2 doctoral students' perceptions of identity construction in their academic writing, this study can promote L2 writers' awareness in terms of how to project an effective and authorial self in writing, and thus improve the credibility of their writing and their images as competent academic writers.

1.6 Organization of the dissertation

The dissertation is organized as follows. In this chapter, I have discussed the current problems with regard to writer identity construction among L2 doctoral students in the ESL settings, and the significance and purpose of the study.

In chapter two, I provide an overview of relevant literature, concentrating on the theoretical conceptualizations of writer identity, the empirical research focusing on the construction of identity through different sets of textual features, and the factors that can affect the construction of identity. I also review prior studies related to L2 doctoral students' identity construction in written discourse, and discuss the research gaps that still exist. This chapter is ended by putting forward the research questions.

Chapter three concerns the methodology adopted in this study. I start by presenting the rationales of adopting the qualitative case study design, and then discuss the context that this study is situated in, the rationales for selecting my participants, followed by the profiles of the three participants examined in this dissertation project. I then present the data collection and analysis procedures. I also discuss my subjectivity as a researcher in this qualitative multi-case study and the validity of this study.

In chapter four, I report the findings of how identities were constructed through metadiscourse in the three participants' papers, the patterns in the metadiscourse usage to construct identities emergent from the papers, and how identities were constructed under the influence of *autobiographical self*, *possibilities for selfhood*, and agency. After reporting the results of each case, I make a cross-case comparison and highlight the similarities and differences found in identity construction among my participants.

Chapter five covers the discussion of the results of the three cases, the insights for doctoral academic writing and identity construction, and the limitations of the dissertation study.

CHAPTER 2: LITERATURE REVIEW

This chapter covers the relevant literature concerning L2 writer identity construction. The literature will be demonstrated in the following way: First, I provide an overview of the theoretical orientations of identity, and specify the theoretical underpinning of this study is social constructivism, which underlies the subsequent discussion of the conceptualizations of writer identity. Next, I review empirical studies related to L2 writer identity from different perspectives, including linguistic features that contribute to identity construction and the challenges L2 writers face during the process of constructing identity. Then I elaborate on the current writer identity scholarship related to doctoral students. Last but not least, I discuss the insights that have been gained in current research, address the research gaps that still exist and how this body of research motivates this dissertation study, followed by the research questions this study aims to investigate. The overview of the review of literature is displayed in Figure 1.

Figure 1. Overview of the literature review

2.1 Identity and writer identity

In this section, I start with the introduction of two concepts: identity and writer identity. Beginning from the analysis of identity, a more general concept, I gradually shift the focus of discussion to writer identity. Then I offer an overview of theoretical underpinnings of identity. Finally, the conceptualization of writer identity is presented.

2.1.1 Writer identity being a component of identity

The "basic structure of the idea we have of ourselves and other people" (Elias, 1998, p. 270), i.e., identity, is an "indispensable certainty lodged within language and its conventional usages, without which individual orientation, interpersonal communication and joint action would all be impossible" (Williams, 2000, p. 1). When approaching writer identity, it is essential to first consider the construct of identity as the former is a branch form of the latter: Identity being a more generic and overarching concept, underlying the foundation of more specific and concrete labels of writer identity. In the scenario concerned in this dissertation, writer identity is an aspect of identity in general. As an umbrella term, identity first of all constitutes multiple facets, such as being an English teacher, applied linguist, novelist, or second language writer, to name just a few. In Gee's (2017) term, these are called *activity-based identity*, given the fact that different kinds of identities are constructed out of different circumstances and activities. For example, a teacher constructs their identity in such activities as preparing and delivering classes, interacting with students, and offering oral and written feedback on students' assignments. Thus, to make it specific, identities emergent from different social activities comprise different areas of research; writer identity, which focuses on the identity writers formulate during the act of writing, is one of them. Second, one individual can be characterized by potentially numerous identities simultaneously, such as being a doctoral student, an L2 writer, a talk show fan, a documentary lover, a cat person, a fanfiction writer, a heavy metal music lover, and so on, and these identities are not in conflict with each other. Instead, within the entity of one individual, these subtle, blended, or even blurred identities are perfectly compatible (Ochs, 1993). They are triggered and become prominent under different occasions, and in turn, these branch identities altogether contribute to a collective identity, i.e., who the person is. In this sense, one's writer identity branch is shaped by one's general identity; it can also contribute to the overall manifestation of who the person is.

In light of the above two reasons, writer identity and identity are inextricably related to each other. Therefore, before I offer a definition of writer identity from different perspectives, it is important to first review research focusing on the theoretical conceptualizations of identity, with the aim of informing and mapping onto the analysis of writer identity later in this section.

2.1.2 Theoretical orientations of identity

Identity has been approached from various disciplines such as art, history, psychology, ethnography, and sociology (e.g., see Bosma et al., 1994), and different disciplines have different foci, and rationales behind the conceptualizations of this construct. In this section, I review different theoretical orientations of identity from essentialist, social constructionist, and social constructivist perspectives.

2.1.2.1 Essentialist and social constructionist views of identity

The cultural essentialist view of identity refers to "discourses and practices which label and relate to particular groups of people in ways which suppress differences and homogenize and fix them, not merely stereotyping but either pathologizing or wrongly idealizing them" (Sayer, 1997, p. 454). As a philosophical concept, essentialist views identity as a fixed, stable, deterministic, and homogeneous construct. However, identity is not "socially determined but socially constructed" (Ivanič, 1998, p. 12). Postmodernist account treats identity as characterized by multiplicity and fluidity, rather than a fixed and monolithic entity. It holds that identity is socially constructed and contingent upon local contexts (Ivanič, 1998, p. 309; Ricento, 2005). Being a term that "silences or short-circuits argument," essentialism in social science literature has been associated with "overwhelmingly derogatory" (Sayer, 1997, p. 453) connotations.

Essentialism has often been countered by social constructionism. As a sociological concept, social constructionism looks at identity as an emergent and socially constructed practice. According to sociologist Peter Berger (1963), "[i]dentity is socially bestowed, social sustained and socially transformed" (p. 98). It is considered that identities, knowledge, and values are not static or deterministic; instead, they are in a state of change, which undergoes negotiations, contestations, and transformations on an ongoing basis realized by language and other semiotic systems (Duff & Uchida, 1997; Hall, 1995; He, 1995). This conceptualization of identity as a construct that is “a process of continual emerging and becoming” (He, 1995, p. 216) also highlights the importance of context in defining identity, as any social activities are situated in contexts of different kinds, which change constantly and engage in the consistent process of contextualizing (du Gay, Evans, & Redman, 2000; Ivanič, 2006).

Highlighting the close relationship between individuals and context, Norton (1997, p. 410) defines identity as “how people understand their relationship to the world, how that relationship is constructed across time and space, and how people understand their possibilities for the future.” An alternative understanding of identity admits people’s subjectivity and the co-construction of identity between individuals and the counterpart with whom they interact. Gee (2000) conceptualizes identity as “being recognized as a certain ‘kind of person’, in a given context”, and the multiple identities constructed by humans connect not only to their “internal states” but also “their performance in society” (Gee, 2000, p. 99).

Instead of just focusing on language related product which helps position one’s identity, Benwell and Stokoe’s (2006) definition broadens the scope of identity construction beyond physical entities embodied by semiotic signs:

> [Identity is] a public phenomenon, a performance or construction that is interpreted by other people. This construction takes place in discourse and other social and embodied conduct, such as how we move, where we are, what we wear, how we talk, and so on (p. 3).

Seeing from a sociolinguistics perspective, Bucholtz and Hall (2005) offer a broad and open-ended definition of identity: It refers to “the social positioning of self and other” (p. 586). They also argued that in sociolinguistics, identity is a linguistic phenomenon, which is not realized by single linguistic devices but is formulated at multiple levels simultaneously. They claim it is during interactions that semiotic signs are endowed with meaning, and propose that identity is discursively constructed and emerges in interaction.

2.1.2.2 Social constructivist view of identity

Social constructivism, as a nomenclature that is similar to social constructionism, bears some subtle differences with the latter. Therefore, some explanation and distinction are warranted. Despite the differences, there are some common ground shared by these two theoretical orientations: They both acknowledge the socially constructed nature of identity. Social constructionism emphasizes the role that social contexts play in shaping and constructing identity (e.g., Matsuda, 2015; Prior, 2001). However, it does not emphasize individual agency in the process of identity formation. Social constructionism denies the possibility for individuals to directly derive meaning from social interactions and contextualized situations (Kim, 2001). In social constructivist view, communities and contexts play an indispensable role in creating identity; individuals, on the other hand, have agency to select from the available repertoire and identify with some but possibly reject others. To use Matsuda's (2015) words,

> the social-constructivist view of voice takes the formation of social conventions into consideration; yet, unlike the social-constructionist perspective, the social constructivist perspective focuses not only on how social norms arise and become stabilized but also on how individuals shape the form and meaning by using the tools provided by the norms – or socially available discursive repertoire. (p. 149)

Largely overlooking individuals' subjectivity and agency, the theoretical orientation of social constructionism unavoidably would generate an assumption that a uniform identity can be ascribed to different individuals from the same discourse community (Ricento, 2005). Social constructivism, on the other hand, holds a more eclectic view toward identity in that it takes individuals as the "interface," and accentuates both the role of one's pooling available social conventions to build identity and the function of human agency in presenting themselves in social practices. During the act of writing, individuals are also performing writer identity construction, the process of which can be conscious or subconscious; their agency always plays a role throughout the process.

The role that agency plays in social-constructivist view does not solely mean that individuals can decide to choose certain discursive representations over others in language production, it also influences the evolving social conventions and repertoire regarding the identity construction. That is to say, individual is not a passive receiver regarding the kinds of identities they should take on; instead, individual identity can shape, and influence the future structure of identity formulation. This dialectical relationship can be summarized in Berger and Luckmann's (1966) words,

> Identity is formed by social processes. Once crystallized, it is maintained, modified, or even reshaped by social relations. The social processes involved in both the formation and the maintenance of identity are determined by the social structure. Conversely, the identities produced by the interplay of organism, individual consciousness and social structure react upon the given social structure, maintaining it, modifying it, or even reshaping it. (p. 194)

Viewing identity as a product both situated in sociocultural contexts and selectively appropriated by individuals in written discourse, social constructivism is the theoretical basis for this dissertation.

While different theoretical orientations perceive identity through different lenses, it may well demonstrate the attention and importance of identity scholarship in the academia across various disciplines. Somewhat contradictorily, the pervasiveness of identity makes it hard to provide a general definition. As Michael (1996) puts it, "it is the very definitional amorphousness that gives the concept of identity its resonance" (p. 7). In the previous interpretations of identity from social constructionist and social constructivist orientations, there are several points that are prominent in order to understand this construct. First, identity is not a unitary or monolithic entity; on the contrary, it is context-sensitive and undergoes a constant process of changing, negotiation, contestation, reformulation, and reconstruction. Second, the construction of identity is controlled both by contextual factors such as one's social and disciplinary background, conventions from one's discourse community, and individual factors such as one's agency (Giddens, 1991). However, the degree to which one can exert their agency is often not infinite: It is constrained by social and contextual factors.

2.1.3 Writer identity: linguistic resources mediated by interactions between writer and reader

Viewed from a sociocultural perspective, academic writing can be seen as a social practice entailing multilayered interactions with the academic discourse community, during which process writer identity can be constructed. Writer identity, in the simplest of terms, refers to the image of the writer created through and by written texts. As Ivanič (1998) notes, writer's identity strongly pertains to individuals' writing: "Writing is an act of identity in which people align themselves with socio-culturally shaped subject possibilities for self-hood, playing their part in reproducing or challenging dominant practices and discourses, and the values, beliefs and interests which they embody" (p. 32). As writing and writer identity are inseparable, there are three constituents indispensable in the construction of writer identity: writer, text, and readers. Writer is the active agent that creates texts, and readers perceive how writers construct themselves through texts which is mediated by semiotic manifestations the writer consciously or subconsciously chooses. In the social constructivist view, the role of language is "that of mediator between the learner and the world, shaping and extending thought" (Hirtle, 1996, p. 91). As noted by Silva and Matsuda (2002, p. 253), "the writer's task is not as simple as constructing an accurate representation of reality; the writer also has to negotiate, through the construction of the text, his or her own view of these elements of writing with the views held by the readers". The linguistic choices writers make have to draw in, influence and persuade readers (Hyland, 2002a). In order to achieve the persuasive goals in (most) academic genres (Hyland, 2004a), writers are expected to address the following communicative purposes (Hyland, 2004a, p. 12):

- establish the novelty of one's position
- make a suitable level of claim
- acknowledge prior work and situate claims in a disciplinary context
- offer warrants for one's view based on community-specific arguments and procedures
- demonstrate an appropriate disciplinary ethos and willingness to negotiate with peers.

As writing research has long ago shifted its attention from writing as a product to writing as a process, what writers are doing during writing in order to attain the effect that they intend their writing to achieve had caught the researchers' eye. However, this body of research concentrated on "what the writer is doing, not what s/he is being" (Ivanič, 1998, p.94). As a relatively long-lasting medium through which knowledge can be displayed and constructed, writing can demonstrate the writer's identity: It can reveal their personality, dispositions, values, background, interest, scope of their reading, their field knowledge, discourse communities they belong to, the list can go on. Writers do not only perform the social act of writing; they write themselves into the text (Clarks & Ivanič, 1997). Hyland (2010) suggests the regular and patterned use of linguistic features can help create a coherent and stable writer identity. The impressions conveyed from a text is important both for the writer and their intended audiences. Usually, academic writers would hope to create an authorial and reliable impression through writing. However, there are sometimes mismatches between what the writer (consciously or unconsciously) intends to convey of themselves and what the readers truly perceive (Burgess & Ivanič, 2010), as the disciplinary background, field knowledge, prior learning experience, interest, and values among writers and readers are not necessarily similar. As a result, the intended effect given off by a piece of text can vary among different readers. A piece of written work, upon completion, is open for readers' interpretations, explanations, and understandings based on their distinctive background. Thus, in defining writer identity, audience is an indispensable variable in completing the process of the positionality of oneself (Matsuda, 2015).

It is obvious that the writer is the active agent that patches a piece of text: Without the writing act of an individual, a text will not be produced. However, the writer is not the only agent that plays a part in the construction of writer identity: It is also "formed by others" (Frick & Brodin, 2020, p. 211). It is negotiated by readers' perceptions and interpretations (e.g., Tardy & Matsuda, 2009) based on their prior experience, ways of approaching knowledge, and discourse community conventions that they belong to. It can be said that upon the completion of a piece of writing by the writer, identity construction is still under the process of further completion, which is accomplished by readers, who endow and bring more meanings to the text. Identity construction in one piece of written text can be perceived differently by different audience; it is also possible that the same reader perceives the same piece of text differently across timescales. Therefore, it can be claimed that identity is not a fixed and homogeneous entity; it is contingent upon its immediate interactional contexts and changes overtime.

Nowadays identity tends to be viewed as a dynamic construct, which can be used as an analytical tool to approach schools and society (Gee, 2000). There is no single piece of text that does not demonstrate the being of the writer: Identity is not optional (Clark & Ivanič, 1997; Matsuda, 2015). Even in academic written discourse, which had been generally considered as detached, impersonal, and devoid of human voice, identity is unavoidably embodied through words. In academic discourse, “writer's voice is always embedded in ways of knowledge making and writing traditions” (Stock & Eik-Nes, 2016, p. 89). Academic writing is conceptualized as a “persuasive endeavor involving interactions between writers and readers” (Hyland, 2005, p. 173). During writing, writers perform interactions of different kinds: They constantly have conversations with their potential readers; by situating their writing in existent scholarship, they have interactions with previous texts, the one they are writing at the moment and future voices; During the process, writers also exercise their agency to conform to particular values, conventions, rules, and beliefs (Ivanič, 1998).

Shaped by multiple factors, writer identity “emerging from the text is partly the responsibility of the writer, partly the responsibility of the reader, and partly the responsibility of the socio-cultural context which supports the discourses they are drawing on” (Ivanič, 1994, p. 6). The ways a writer constructs a piece of academic writing are shaped as well as constrained by social conventions, which can be influenced by macro sociocultural systems as well as micro discourse community conventions. Writing is not an artefact that is produced in a vacuum; instead, it is a constellation of the writer’s perceptions shaped by the communities they are associated with; meanwhile the very act of constructing writer identity performs a sociohistorical role by contributing to the construction of the discourse communities in the long term. There is "a dual relation between identities and the landscape of practice: they reflect each other and they shape each other" (Wenger, 1999, p. 190).

In terms of its definition, identity, or self-representation in written discourse, has been conceptualized along different dimensions in previous literature. Cherry (1998) approaches self-representation through the differentiation of *ethos* and *persona* in written discourse, with *ethos* referring to "writer's 'real' self" and *persona* being "audience invoked" and the "writer's 'fictional' self" (p. 399). The author argues that the two terms play complementary roles in facilitating understanding of self-representation in written discourse, though they originate from different traditions. Highlighting the roles of both writers who create texts using discursive and non-discursive devices, and readers who are the agents that perceive and interpret the effect conveyed through written text, Matsuda (2001) defines writer identity as "the amalgamative effect of the use of discursive and non-discursive features that language users choose, deliberately or otherwise, from socially available yet ever-changing repertoire" (p. 40).

2.2 Constructing writer identity: relevant constructs

In this section, I mainly define key terms that are related to the framing of writer identity in academic discourse. Beginning with the discussion on necessary factors that contribute to writer identity formation, I then discuss social constructivism, which lays a theoretical foundation for how this dissertation perceives writer identity. Then I proceed to define two more specific terms that are relevant to writer identity construction in this dissertation: agency, and genre.

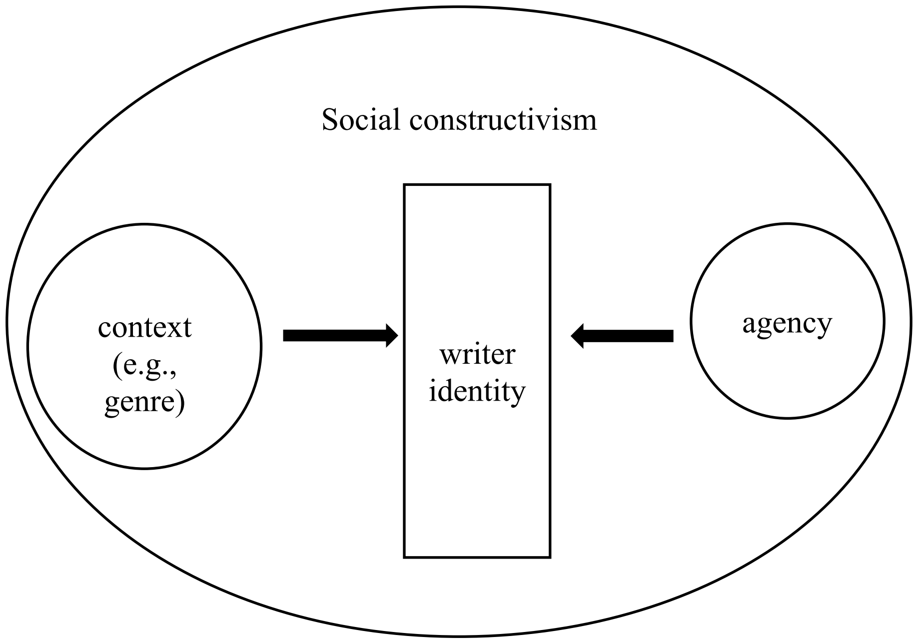

Figure 2. Conceptualization of the construction of writer identity

As has been said, context is a crucial shaping force in the construction of identity; therefore, one question that is worth further probing is, more specifically, what contextual forces potentially contribute to the overall construction of writer identity. Under the overarching theoretical orientation of social constructivism, there are a few factors that come into play collaboratively and simultaneously. According to Adams and Marshall (1996), "[i]dentity is a social-psychological construct that reflects *social influences* through imitation and identification processes and *active self-construction* in the creation of what is important to the self and to others" (p. 433, emphasis in original). From a sociocultural lens, socially available discursive repertoire offers possible linguistic devices from the discourse community from which writers can select to represent their identity in writing. Genre is an important strand within the expansive scope of context that performs an essential role in writer identity construction. From an individual writer's perspective, during writing, they can act proactively either to conform to or resist certain conventions exerted by sociocultural forces. Therefore, concurrently context and agency function together to help writers build their identity during the writing process. The formation of writer identity collectively by social and individual forces is summarized in Figure 2. The more specific ways in which each of these factors operates are discussed in the following sections.

2.2.1 Social constructivism

Identity has been approached from various theoretical lenses such as the sociocultural linguistic approach (e.g., Bucholtz & Hall, 2005) and social constructionism (e.g., Prior, 2001); an important one for examining writer identity is social constructivism (Matsuda, 2015; McKinley, 2015). Social constructivism postulates that learning happens when learners are actively engaged in social activities, and knowledge and understanding of the world is developed by individuals collaboratively during interactions (Amineh & Asl, 2015; Vygotsky, 1978). From a social constructivist view, writing is perceived as a social practice. During writing, individuals can recourse to external resources to back up their arguments, claims, and ideas; at the same time, they will also internally make their own decisions to choose from the socially available repertoire. In other words, writers are informed and motivated by sociocultural conditions during writing, yet they simultaneously exert agency to select the linguistic expressions, which eventually could influence the evolving conventions and norms in the social world. In Berger's (1963) book, an institution, "a distinctive complex of social actions," is conceived of "as a regulatory agency, channeling human actions in much the same way as instincts channel animal behavior" (p. 87). Acknowledging both roles of social conventions and individuals' agency in meaningful learning and identity building, social constructivists view writer identity as shaped by contextual artefacts and "how the social conventions are appropriated by individual writers as they respond to the particular rhetorical situation in the process of writing" (Matsuda, 2015, p. 149). The construction of identities in written discourse also entails a process where individuals contribute to the gradual shifting, reconfiguration, and fortification of identity convention particular to a specific discourse. In other words, it is characterized by recursiveness and mutuality between individual writers and contexts: socially available linguistic repertoire exerts shaping and restrictive forces upon individuals' writing practices, while individuals' writing practices can "project into the future, contributing to the reshaping of socioculturally available practices and communicative resources for future events in the same or similar contexts" (Ivanič, 2006, p. 10). This orientation of theorizing writer identity corresponds to the conceptual framework of writer identity in this dissertation, which will be expounded closely later in this chapter.

2.2.2 Agency

Though writer identity is socially constructed, it does not mean that writers slot into the predetermined identities with the social practices they have already acquired; instead, individual writers could always negotiate depending on their personal background and intentions (Hyland, 2002a). According to Clarke (2009), engaging in identity work is immanently connected to exercising one's agency. From a social constructivist approach, in a sociocultural setting, there are many, yet limited, identities available; but each individual's manifestation of their identity in text is unique: They have "the freedom to select from the culturally available voice types and/or creatively recombine them in their own way", and "can exercise the power to conform to or resist the social forces that are privileging one voice type over another" (Ivanič & Camps, 2001, p. 7). To look at identity from an agential perspective means to attend to how individuals construct identities instead of how individuals conform to identities socially imposed on them (Shanahan, 2009).

The construction of identity coexists with individuals' agency (Norton & Gao, 2008). For example, in writing an academic paper, writers are strategic planners: They tailor their communication for particular settings (Tracy & Robles, 2013), or they could possibly struggle to reconcile the voice of their own and expectations from their disciplines (Aitchison et al., 2012). Butler (1997) argues that identity is not an a priori notion but is created through social practices where "the subject emerges both as the *effect* of a prior power and as the *condition of possibility* for a radically conditioned form of agency" (p. 14-15, emphasis in original). Ahearn (2001) defines agency as "the socioculturally mediated capacity to act", and it is emergent "from the social, political, and cultural dynamics of a specific place and time" (pp. 112-113). Accentuating the dialectical relationship between agency and context, "writing-related agency is influenced by the multiple relationships, resources, practices and experiences" available to writers (Botelho de Magalhães et al., 2019, p. 7), and writers' "agencies and investments in language learning and use are shaped by the range of identities available for them in the L2" (Pavlenko, 2001, p. 319).

Exploring writer identity from a social-constructivist view encourages us to mull over the fact that individual writers are immersed in complex social structures (Botelho de Magalhães et al., 2019) where one's agency is bi-directional (de Fina, Schiffrin, & Bamberg, 2006, p. 2). One the one hand, sociocultural forces position who the writers are and situate individuals in dominant discourse and practices. On the other, individuals can use their subjectivity and reflexivity in making choices to negotiate and perform identities vis-a`-vis others and the dominant discourse and practices. As claimed by Sewell (1992), "Structures shape people's practices, but it is also people's practices that constitute (and reproduce) structures. In this view of things, human agency and structure, far from being opposed, in fact presuppose each other" (p. 4). Agency is inscribed in writing practice through writers' interactions with different entities at varying levels: Readers constantly have conversations, either consciously or subconsciously, with their imaginary but targeted readers, and they have conversations with previous texts, the one they are writing at the moment, and future voices. This is also a process where writers choose to adopt certain values, beliefs, and practices while rejecting others (Ivanič, 1998).

2.2.3 Genre

Genre, according to Hyland (2015), refers to "abstract, socially recognized ways of using language" (p. 32). Genre as a construct that can mark the boundaries for writers to "structure[s] the possibilities available to the individual" (Ivanič and Camps, 2001, p. 6) and shape writers' discursive choices and their construction of identities. However, the concept of genre can cause some confusion (Fairclough, 2003) because on the one hand it can be a broad term referring to highly abstract categories of discourse, on the other, it can be situated within concrete social practices signaling a specific type of text. In other words, the degree of abstraction is associated with different types of genres, and "genres vary in terms of the nature of the activity they constitute or are a part of" (Fairclough, 2003, p. 73). For example, academic writing, if compared to fiction, is a type of genre on a high level of abstraction, while dissertation is a more particular form of genre of "academic writing."

Swales (1990) has encouraged us to see genre in a contextualized lens, i.e., in the discourse community where it is used. Fairclough (2003) notes that genre is "defined in terms of the purposes of the activity" (p. 70). And it is common that there are multiple purposes that a text aims to attain. For academic writing (here I am on an abstract level), its purposes are varied depending on the types of genres, but there are several that are commonly distributed, such as displaying and constructing knowledge, and conveying a credible writer image through the text, no matter what the academic genre specifically is. As Bhatia (2004) stated, genres may have common communicative purposes, “most of them will be different in a number of other respects, such as their disciplinary and professional affiliations, contexts of use and exploitations, participant relationships, audience constraints and so on” (p. 57). Given this, variations in regard to the linguistic uses as well as identity construction across genres can emerge in accordance to contexts in which the genres are embedded.

Text is produced by individuals who act upon choices and constraints available to them (Hyland, 2015). This makes clear that, to some extent, as a “framing device” (Paré, 2019, p. 2), genre has a prescriptive function in terms of guiding writers to write in ways acceptable to a specific community of practice. Assuming a framing role in shaping the regularities that writers are expected to conform to, genres are dialectically stable and flexible at the same time (e.g., Worden, 2018). The recurrent communicative situations call for patterned discursive conventions; however, writers can choose from the rhetorical strategies yet possibly resist others when responding to specific rhetorical situations. It is worth pointing out that the restraints exercised by genres are not only from academic communities. From a micro perspective, the institutions, academic programs, and course guidelines can all stipulate rules, though some of them can be tacit, nuncupative and nuanced, for writers to follow.

Genres “constitute particular social relations between interactants” (Fairclough, 2003, p. 75): Usually genres also presuppose the relationships between the writer and their target audience who are the evaluators, readers, or reviewers of the paper, which will influence the way writers present themselves. According to Bakhtin (1986), “both the composition and, particularly, the style of the utterance depend on those to whom the utterance is addressed, how the speaker (or writer) senses and imagines his addressees, and the force of their effect on the utterance (p. 95)". Readers, who interpret texts, also play a role in shaping the quality of writing. For example, when discussing the role of teachers as the audiences of students' writing, Li (1992) points out that teachers are not "innocent readers applying some universally agreed upon standards in our evaluation of student papers" (p. 5), and "the goodness of writing does not reside wholly in student texts; it also resides with teachers who judge and grade their papers" (p. 4).

What is worth mentioning is that currently, audience is not viewed as a stable and unitary entity; it is "hardly a homogeneous grouping" (Hyland, 2001, p. 551). Griffin (2006), for example, conducted a study with two undergraduates in a British university, and analyzed their teacher-assessed project report along with other supporting materials such as textbooks, course materials, and so on. The study found that audience is a fluid and variant entity; it varies and changes within a text, just like, correspondingly, identity undergoes constant negotiation and fluctuation in a text.

Genres can be a significant variable in shaping identity construction because the available pool for writers to construct themselves as legitimate authors of their writing varies from genre to genre. For instance, student writers, when writing a final paper for a course, usually try to thoroughly demonstrate relevant field knowledge that is fundamental to the topic in order to construct themselves as hardworking and conscientious students who have done enough preparation. However, the same writer would write differently when their aim is to write a paper for publication where more general but focused background knowledge would be a preferred rhetorical strategy to claim authority. Therefore, genre is a variable that impacts writer identity formation because it has innate, tacit, as well as explicit and sometimes implicit discoursal conventions that writers are expected to follow (e.g., Swales, 2014). To use Ivanič and Weldon's words (2014) in their descriptions of the role of discourse and genre play in the formation of writer identity: "Discourses carry possibilities for self-hood in terms of the values and beliefs to which writers who deploy them become party; genres carry possibilities for self-hood in terms of the social relations and communicative purposes to which writers who deploy them become party" (p. 169).

In a nutshell, self is formed by the organic synthesis of both internal self-definition, which could be embodied by one's agency, choice and preferences, and external variables, manifested in sociocultural, academic, disciplinary and institutional conventions (Jenkins, 2008). Hyland's (2012) words can well demonstrate how writer identity is constructed in a context and simultaneously shaped by personal agency: "We draw on a repertoire of voices as we communicate, bringing to the task our own experiences, purposes and conceptions of self to recombine the options offered by the genre to perform a community identity" (p. 14).

This section has reviewed relevant constructs that collectively function together to influence the construction of writer identities. From a social constructivist perspective, identity is not a fixed, static, or stable construct. Instead, it is fluid, ever-changing; it is also contested, and negotiated under various factors, such as context, genre, and personal agency, and undergoes gradual change overtime (Burgess & Ivanič, 2010). The variables concerned in this section play important roles in shaping one's identity in written texts. That is, under the overarching framework of social constructivism, the external constraining factors, such as genre and the internal shaping forces, such as agency, work in tandem with each other to realize the enactment of writer identity.

2.3 Theoretical conceptualizations of writer identity

As Bucholtz and Hall (2005) points out, the theoretical approaches to identity have been of secondary concern because of lavish attention having been paid to linguistic research on identity construction. For this reason, theoretical conceptualizations of writer identity are rather scarce in current academia. Among the theoretical conceptualizations of writer identity, Ivanič's (1998) study based on mature writers is one of the most prominent works so far. In her (1998) seminal work on writer identity, writer identity refers to "aspects of the identity of an actual writer writing a particular text" or "abstract, prototypical identities available in the socio-cultural context of writing" (p. 23). The former one consists of three strands, namely, "*autobiographical self*", "*discoursal self*", "*self as author/authorial self*", while the latter refers to "*possibilities for self-hood*" (p. 23).

As stated by Benwell and Stokoe (2006), "[A]lthough people may present themselves differently in different contexts, underneath that presentation lurks a private, *pre-discursive* and stable identity" (p. 3, emphasis in original). This facet of identity is somewhat corresponding to *autobiographical self* in Ivanič's framework. *Autobiographical self* is concerned with "a writer's sense of her roots, of where they are coming from," and it is "socially constructed and constantly changing as a consequence of their developing life-history" (Ivanič, 1998, p. 24). It involves all the life experience that the writer has accumulated up to the moment of writing, with their associated interests, beliefs, and values (Burgess & Ivanič, 2010). *Discoursal self* involves "the impression — often multiple, sometimes contradictory — which they [writers] consciously or unconsciously convey of themselves in a particular written text" (p. 24). Ivanič argues that "this identity is constructed through the discourse characteristics of a text" (p. 24), which are all situated under specific social contexts. In Ivanič's book, she corresponds the two layers of identity to Goffman's distinction between writer-as-performer and writer-as-character respectively. The former refers to the writer who "brings previous life experiences and social positions to the performance of self" (Matsuda, 2015, p. 143). This sense is more relevant to the tangible experience and physical characteristics of the writer, which is one's *autobiographical self*. On the other hand, writer-as-character is "created through the features of written discourse" and "portrayed as a result of the writer's performance" (Matsuda, 2015, p. 144), which correspondingly is the writer's *discoursal self* and *authorial self*, as discussed next.

Self as author or *authorial self*, "as an aspect of discoursal self," also concerns the identity constructed by and through the texts; it refers to the writer's sense of authority in their writing. It is a "relative concept" (Ivanič, 1998, p. 26), as the degree of the presentation of authoritativeness is not categorical, but highly dependent on readers' perception: Some may be authoritative in texts while others sound more deferential and cautious, trying to conceal themselves behind writing. The degree of authoritativeness is relative. It is of particular importance in academic discourse (Ivanič, 1998), as it characterizes an individual's right, confidence, and qualifications to claim membership in certain discourse communities (Ouellette, 2008). Writers might differ considerably in presenting *self as author*: Some might be strong in stating their argument and be authoritative, while others are uncertain, hedging to varying degrees. The different degrees of demonstrating authoritativeness will have an impact on readers' different perceptions of writer's self as author. However, the same piece of text demonstrating equal degree of *self as author* can be perceived differently by different readers: The text to someone that is authoritative and scholarly might seem arrogant and/or generalized to another reader.

The three aspects of writer identity all concern the actual writing activities, because "a subject's 'being' is inextricable from their 'doing'" (Ivanič, 2006, p. 11). However, though there is a clear distinction among the three, they are "neither completely discrete nor mutually exclusive" (Matsuda, 2015, p. 144). As Kress (1996) puts it, writings produced are "saturated with the meanings of their makers in every aspect of their form" (quoted from Scott, 1999, p. 180). In this sense, people's "doing" is inseparable from their "being": Their being, also known as *autobiographical self*, serves as the foundation based on which they are entitled to the eligibility, choices, and even the limitations to accomplish their doing. The *autobiographical self* is the materialistic background, helping construct *discoursal self* and *self as author*, while *discoursal self* and *self as author* are two sides of a coin, contributing to a complete picture of writer identity construction co-built by writer and readers and mediated by discursive and nondiscursive features in text. Yet *discoursal self* and *self as author* will also bring about the emergence of *autobiographical self* through texts.

These three aspects of writer identity interweave together also in that they are "both enabled and constrained by the *possibilities for self-hood*" (Matsuda, 2015, p. 144). This is the abstract dimension of the conceptualization of writer identity. *Possibilities for self-hood* is concerned with "prototypical possibilities for self-hood which are available to writers in the social context of writing: 'social' identities in the sense that they do not just belong to particular individuals" (Ivanič, 1998, p. 27). According to Block (2007), "although identity is conditioned by social interaction and social structures, it conditions social interaction and social structures at the same time. It is, in short, constitutive of and constituted by the social environment" (p. 865-866). When writers are performing the act of writing, the choices they can make to construct themselves is not a pool of infinite possibilities. Instead, there is a limited repertoire that they can choose from, which shapes and constrains the identities they create through the text. Figure 3 displays the relationships and interactions of the four strands of writer identity in Ivanič's (1998) conceptual model.

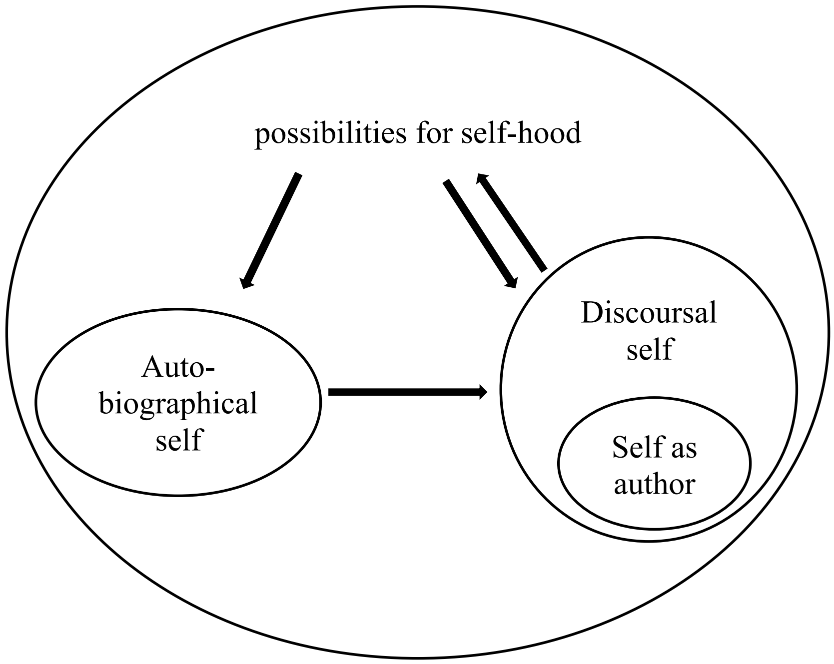

Figure 3. Four strands of writer identity in Ivanič's (1998) conceptual framework

Ivanič's (1998) conceptual framework of writer identity will be the guiding framework in this study. Informed by the social constructivist approach to identity, this study treats equally the roles of contextual factors and individual agency in the construction of writer identity. In Ivanič's (1998) model, writer identity is a construct that is framed in situational contexts, while writers can actively choose from the socially available repertoire, such as preferred writing conventions in a specific genres, to construct their identity. The contextual and individual factors are essential in Ivanič's (1998) framework under social constructivist approach. The interactive, related, and dynamic relationship between social constructivist approach and Ivanič's (1998) framework is shown in Figure 4.

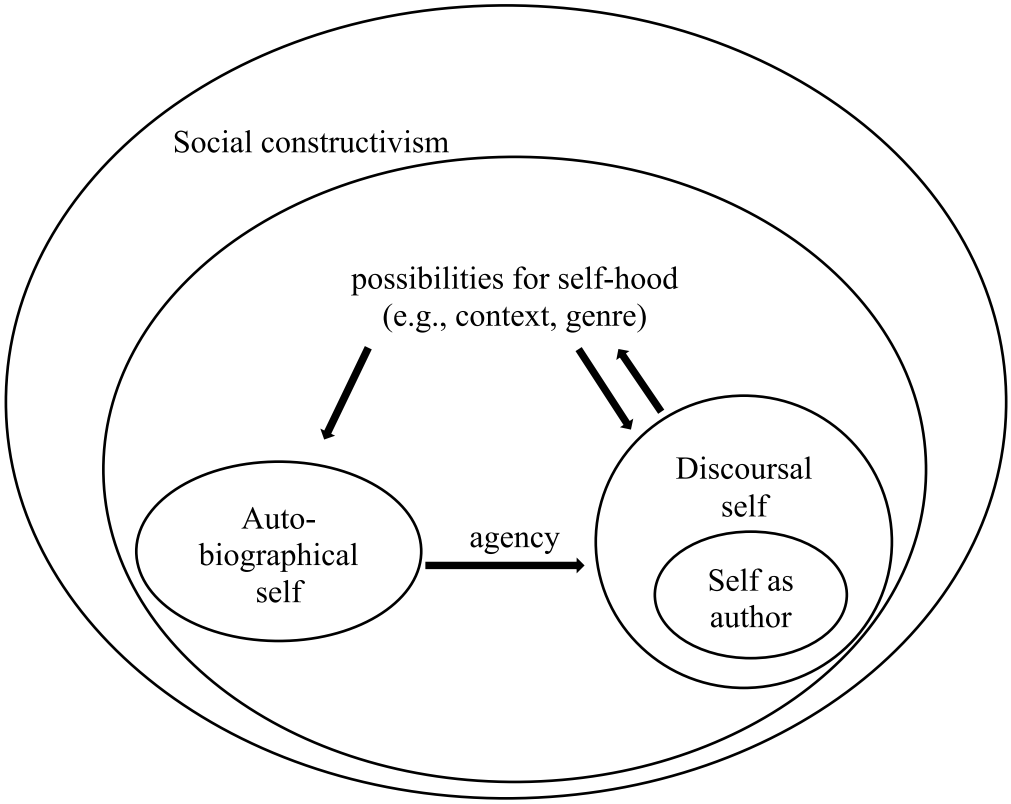

Figure 4. Conceptualization of construction of writer identity under social constructivist approach

Ivanič's (1998) framework of writer identity has been employed in current identity scholarship by quite a few studies investigating L2 writers. For example, with Ivanic's framework as the analytical focus, Alkhanbooli (2018) did a qualitative study on three undergraduate/graduate Arab students studying in the US and found that the students tended to transfer their L1 writing skills and strategies when writing in English. It concluded that their writer identity in English writing was similar to what they embraced in Saudi Arabia, i.e., they constructed the same *autobiographical self* in both languages. In the meantime, despite the struggles they had in writing in L2, they still managed to reveal themselves as knowledgeable intertextual writers, reader-concerned writers, reliable writers, flexible writers, and so on.

Based on Ivanic's framework of writer identity construction, Li and Deng (2019) did a case study on a Chinese undergraduate's personal statement writings over a period of three years, through examining rhetorical moves in her writings. In order to construct a desired identity as an eligible applicant, discoursal moves including narrating life experiences, referring to oneself and others, and highlighting could be found in the student's personal statements. Using Ivanic's updated analytical framework combining writer identity and temporal dimensions (Burgess & Ivanic, 2010), Burgess (2012) conducted a case study with an adult writer and explored how her identity was constructed within a piece of writing about China. Her study highlighted the importance of time in the shaping of writer identity: a single piece of writing involved coordination of processes occurring on different time scales, and the identity projected in writing temporally linked the present, past, and future.

Using Ivanič's (1998) framework on writer identity, Kim, Baba, and Cumming (2006) conducted a multi-case study over three years with three east Asian students (two Japanese and one Korean) studying in an ESL setting, with students' writing samples and interviews as their sources of data. They aimed to find out how the three students represented themselves in terms of the four strands of writer identity in their texts and interviews and how their communities had influenced their identity construction. Through analyzing the linguistic, syntactic, and rhetorical features in the writing samples and drawing on interview data, the study found one of the students revealed her identity by considering the nature of the course. For example, she expressed her *autobiographical self* more vividly in the essays she wrote in the ESL program than in those she wrote for university courses. What's more, there was a qualitative difference regarding the conveyance of *discoursal self* between her writing in the pre-university ESL program and her later academic courses, as she tried to build an identity of belonging to the academic community in academic writing by citing publications, which was not a characteristic in her writings in the ESL program. In the meantime, this student frequently used first person singular pronoun to claim her *self as author*. However, the second student intentionally hid her identity in her essays, and it was easy to tell that she strictly conformed to academic conventions, as she wrote in an impersonal and detached tone. For example, no *autobiographical self* could be found in her essays, and she barely used self-mentions in her writing. The third student exhibited her *autobiographical self* in some pre-university essays, and she strategically used such devices as categorical present tense verbs to convey objectivity in those essays. This student also demonstrated her *discoursal self* in university essays different from in pre-university essays.

2.4 Academic writing and identity construction in interaction

As has been discussed in previous sections, identity is not static or monolithic (Ivanic, 1998); instead, it undergoes constant changes and negotiations under certain context, and is formed *between* individuals and *within* social relations (Hyland, 2012a). In other words, identity is networked through participation in a range of disciplinary practices in writer's community of practice. Writer identity is created through interaction with disciplinary members, which is mediated through text. Through engaging oneself in social relations, one constructs themselves in their encounters with other people in the world (Cameron, 2001), "so… *who we are,* or rather *who we present ourselves to be,* is an outcome of how we routinely and repeatedly engage in interactions with others on an everyday basis" (Hyland, 2012, p. 3, emphasis in original). For example, doctoral students, in order to assume a scholarly identity, have to go through the socialization process where interactions with established figures of the discourse and practices of the community of practice (i.e., advisors, committee members), as well as peers are indispensable social interactions that are always a part of their doctoral journey. Besides the interaction with people from the same institutions, they may also need to branch out to other areas of academia, such as presenting at conferences, and publishing their research to join disciplinary conversations. This way, they can gradually construct their identity as disciplinary insiders and competent scholarly writers.

On a textual level, interaction also exists to shape doctoral writers' identity construction. Writers do not write in a vacuum; for doctoral writers, completing the writing product to address community expectations, which can be manifested through anticipating readers' reactions, is essentially the task they need to undertake, as said by Hyland:

> we do not just say what we think and get over with it, but take care to design a text for particular recipients so that ... it meets the rhetorical expectations, processing abilities and information needs of readers .... The arguments we make, the positions we take and the ways we try to connect and fit in with others, all contribute to the presentation of ourselves and so influence how others respond to us. (Hyland et al., 2012, p. 135).

2.5 Second language learners' writer identity in research

In the past couple of decades, the nature of academic writing has been viewed as a social practice involving writers' interactions with texts and people from their community of practice (e.g., Fairclough, 1992; Wenger, 1999). As a result, writer identity has become an important research focus in academic writing scholarship. For many L2 writers, identity might still be a relatively new concept. While it is the writers' discursive maturity that characterizes their ability to manipulate textual resources to construct a proper identity, L2 writer identity research does highlight the crucial role that identity plays in impacting the outcome of a piece of writing; the mastery of projecting an authorial identity can probably be seen as one of the indicators of discursive maturity (Thorton, 2003). Nowadays, L2 writers, who as a group have been increasing in number, have been one of the foci of current identity research. In this section, I first discuss briefly why identity construction is a challenge in L2 writing, then I review relevant literature with a focus on discursive features that embody writer identity.

2.5.1 Writer identity being a challenge for L2 writers

The sociocultural perspective treats writing as a social practice, which is situated in particular sociocultural context and achieves corresponding communicative functions. In order to realize the rhetorical purposes, one of the challenges student writers face is to project an authorial identity and to claim authorial stance towards the propositional content. This challenge is shaped by various factors. This section provides an overview of the way writer identity constitutes a difficulty for L2 writers.

Stetsenko and Arievitch (1997) noted that the analysis of identity construction does not solely rely on discourse-based approach in that it will neglect the writer's psychological processes. The social constructivist approach perceives that writer identity is inseparably bound to plentiful contextual factors, such as culture, language proficiency, familiarity with a certain discourse community, one's prior life experience, the list can go on. It is very hard to disentangle identity from these external factors, i.e., one's *autobiographical self*, and *possibilities for selfhood*, and solely investigate identity through texts.

There is no denying that discourses and identity construction are influenced and formed by ideologies (Schaffner, 1996). Writer identity, therefore, is a concept that varies among different cultural ideologies and values. Successful writing in L2 does not only presuppose correct use of grammar and vocabulary; projecting authorial voice has become an essential component in the assessment of writing (e.g., Zhao & Llosa, 2008). However, writer identity may have not been a popular notion for L2 writers, particularly L2 writers from a collectivistic cultural background. Constructing a proper identity through writing seems a challenging task for some L2 writers due to multiple reasons. First, for less proficient writers, their attention is more often than not exclusively directed to linguistic accuracy; constructing a writer identity, a more advanced requirement cognitively, socially, and linguistically, can be overwhelmingly demanding for them. Moreover, this may not be a phenomenon unique to writers with insufficient language proficiency. For advanced language learners, identity construction is still a rhetorical function that constitutes an obstacle. For example, Flowerdew's (2001) study on editors' and reviewers' attitudes toward NNS's manuscripts showed that the lack of authorial voice was a common problem among this group of writers. Second, cultural identity plays a profound role in the shaping of writer identity, and it makes available certain ways of us perceiving ourselves (Hyland, 2012), which is the foundation of our *autobiographical self*. According to Shen's (1989) personal example as a Chinese writer who both was socialized into and engaged himself into the English composition community, writing in English essentially entails disassociation with the ideological and logical identities which are endowed and cultivated in his native Chinese culture: "Both had to be modified or redefined in learning English composition" (p. 459). The ideological identity refers to "the system of values that I acquired (consciously and unconsciously) from my social and cultural background", and logical identity means "the natural (or Oriental) way I organize and express my thoughts in writing" (p. 459).

This discrepancy between Anglophone ideology and other cultures can pose considerable difficulties to L2 writers studying in English speaking context regarding their unawareness of identity construction and the transfer from their native writing conventions and various contextual factors. For example, Abasi et al. (2006) found that five ESL graduate students, two Iranians, one Korean, one Mexican, and one Honduran, expressed an unawareness of their writer identities, and these students did not consciously reflect on the identity they constructed of themselves in writing. When ESL writers are expected to write nativelike, i.e., claiming a more assertive and confident self, it is considered that the transition in identity may take them arduous efforts.

The construction of writer identity is heavily shaped by writers' "life history" (Ivanič, 1998). When student writers come from an academically disadvantaged background, it is harder for them to gain academic success and their writing is inclined to display less authoritativeness. Starfield's (2002) research can well illustrate this. She looked into how students negotiated their writer identity when there were asymmetrical sociopolitical and sociohistorical power relations at play by comparing two essays written for a sociology class by a White English native speaker and a Black ESL student respectively at an English-medium university in South Africa: The one written by the White student scored significantly higher than the one by the Black, who failed this task. The successful student, as revealed in the study, was capable of constructing an authoritative and powerful textual identity, whereas the failed student's essay was characterized by a lack of authority, who struggled to manipulate linguistic resources as he was not able to tentatively voice criticism of the authorities, and most clauses in his essay were categorical with scarce use of modalization or modulation. This study showed that coming from an educationally and socially disadvantaged background, the Black student assumed far less textual resources within an academic discipline compared to the White student, and the identity of success was hard to build for Black students.

2.5.2 Linguistic features constructing writer identity

Though writer identity is not an optional extra in writing, it is as an elusive entity, which can be extremely difficult to define and operationalize (e.g., Zhao, 2012). As writer identity is emergent from written text, linguistic resources play a key role in shaping the identity conveyed through writing (De Fina, Schiffrin, & Bamberg, 2006). As Hyland states, "Our preferred patterns of language, in both writing and speech, index who we are…, making the study of discourse a legitimate means of gaining insights into self-representation" (2012a, p. 3), and "our discoursal choices align us with certain values and beliefs that support particular identities" (2002a, p. 1092). Ricento (2005) notes that language has been posited as a prominent marker of social identity and membership. However, operationalizing identity in written texts is extremely difficult (Matsuda, 2015). As identity is an essential element in academic writing, and it is realized lexically, syntactically, and rhetorically (Ivanič & Camps, 2001) in texts, discursive resources are indispensable contributing factors for understanding how identity is constructed in text (e.g., Hirvela & Belcher, 2001). There have been such analytical frameworks as metadiscourse (Burke, 2010; Castillo-Hajan, Hajan, & Marasigan 2019; Hyland, 2004b; Supasiraprapa & De Costa, 2017; Taj, 2017), interactional model of stance and engagement (e.g., Ouellette, 2008), and Systemic Functional Grammar (Danis, 2022; Ivanič, 1998; Ivanič & Camps, 2001; Starfield, 2002) in the current literature aiming to pin down identity in written discourse. Given its importance, in this section, I review literature related to the textual construction of writer identity from the following strands: first person references, Systemic Functional Grammar, and intertextuality, followed by the framework used in this dissertation that contributes to writer identity construction: metadiscourse.

2.5.2.1 First person references and writer identity

Textual features analysis constitutes a major dimension in approaching writer identity. Even though in the extant literature, studies directly connecting stance and engagement and writer identity are not great in number, an important dimension of stance, that is, self-mentions, has been extensively researched both in scope and amount in association with writer identity construction. This is probably because writer identity is also shaped by contextual factors, which cannot be captured simply from text analysis, while self-mentions are a linguistic phenomenon that can be more easily retrieved, through corpus tools in large body of data. Given its prominent role in constructing writer identity, it merits a separate section.

First person pronouns, which manifest the presence of the author, serves as an important medium among the linguistic devices signaling writer identity. Self-mentions have been investigated in previous studies mainly through corpus tools. For example, Hyland (2012) examined self-person pronouns in texts written by non-Anglophone students and published academic English articles. Overall, as the findings suggested, student writers used first-person pronouns four times less than professional writers did. Focusing on student writers, Tang and John (1999) argued that the first-person pronoun *I* did not perform one uniform function. Instead, heterogenous identities could be constructed from this self-mention device. So they worked out a typology of six different identities that *I* could convey through writing (*I* as representative, guide, architect, recounter of research process, opinion, and originator), followed by examining 27 undergraduate essays using the framework. Employing a corpus tool, Asprillia and Hardjanto

(2020) collected 36 articles from four Scopus-indexed English language journals, with half of the articles written by L2 writers. The research found that in order to claim authorial identity, in RAs written by natives and those by non-natives, “we” was the most often used pronouns, while “the researcher” was the most frequently used to demonstrate authoritativeness.

Danis (2022) examined the construction of disciplinary identity across six disciplines (applied linguistics, biology, history, physics, philosophy, and political science) in a corpus composed of journal articles of academic evaluative genres through the co-occurring of “I” and verb process types. The author argued that the use of different verbs after “I” corresponded to different types of writer identity. The results showed that material and mental processes were the most common types across the six disciplines to construct writer identities, while verbal, material and relational processes were found to use differently to reflect writers’ disciplinary selves: authors from hard fields used verbal and material processes to project a scientist image, but social science authors employed the verbal and relational types to engage in disciplinary conversations to formulate a sayer identity, and they used material processes to construct a professional and intellectual identity.

Işık-Taş (2018) compared the use of first person pronouns in three sets of corpora: published RAs written in Turkish and English by Turkish writers, and published RAs written in English by English native writers. The study found that English RAs by Turkish and English authors employed similar amount of first person pronouns, while in articles in Turkish, the number had a significant decline. The study also looked into the discourse functions of first person pronouns; its results revealed that in general, writers preferred to display their authorial identity through using first person pronouns performing medium-risk functions, including stating a goal/purpose, and explaining a procedure, while locally-oriented papers written in Turkish tended to employ pronouns fulfilling low-risk functions, such as representing a community, and guiding the reader through the text. Also adopting a corpus tool, Charles' (2006) examined eight politics MPhil theses and eight doctoral dissertations from the discipline of Materials. In order to build credibility for their research, professional and student writers both were found to choose to reveal or obscure their identities in texts strategically. The study argued low levels of authorial visibility were not necessarily associated with writers' unwillingness to construct an authoritative identity, which depended upon the rhetorical functions the writer intended to accomplish. For example, when taking a critical stance toward other researchers or evaluating other researchers' work, it was necessary for the writers to obscure their self (i.e., avoiding using personal pronouns such as *I* or *we*) and use more objective linguistic structures instead. Starfield and Ravelli (2006) examined the visual representation of self through the title pages, tables of contents, and introductory chapters of a corpus of 20 PhD theses from the fields of history and sociology. In their study, besides investigating the macrostructure of the theses, they specifically focused on the self-representation through the use of "I", the relatively marked form of presenting oneself (Hyland, 2002a), along with its functions. They found "I" performs four rhetorical functions: "I" as guide or architect, "methodological I", "I" as opinion holder, and "I" as originator.

2.5.2.2 Intertextuality and L2 writer identity

From a sociocultural perspective, academic writing is social, which involves interactions with the current text and future audiences. When presenting their research, writers negotiate with their audiences to gain community acceptance; one of the ways of achieving that rhetorical purpose is through establishing intertextual links between the current texts and previous research. According to Kristeva (1980), texts do not only have a horizontal axis that connects writers and readers, but also a vertical axis that links a text to other texts. Intertextuality is realized by citation in a conspicuous manner in academic discourse. Examining international postgraduate students' experience of establishing an academic writer identity in the UK, Boz (2006) argued the negotiation of writer identity was affected by first-person pronouns, patchwriting, and plagiarism, as proposed in previously reviewed studies. Being a common academic literacy practice, citation, which demonstrates the author's acknowledgement of previous literature, can be an indicator of writer identity: The surface form of a citation is closely related to the affective attitude writers take toward the cited propositions.

Research has shown that the *discoursal self* and *self as author* constructed from the text through the use of citations is shaped by *autobiographical self,* and *possibilities for selfhood.* Lee et al.'s (2018) study can provide an example. Their study looked into L2 students' citation forms and their stance toward the literature cited, and found that these students mainly took a non-committal stance predominantly by acknowledging or distancing themselves from the cited materials. Thus their text exhibited deferential tone and lacked critical evaluation of previous literature.

This close relationship of discoursal identity and *autobiographical self* is obvious in other studies. Ouellette (2008), combining Ivanič's framework of writer identity and Hyland's (2005) stance and engagement model, analyzed an NNES first-year undergraduate's writing drafts who was accused of plagiarism. The author specifically focused on *self as author*. In the student's writing, features such as patchwriting and some mechanical errors marked her writing as devoid of authoritativeness. However, her discourse choices were gradually shaped, and in her third draft, it was obvious that she was able to claim some authority by using certain self-mentions. Besides analyzing the text, the author further approached *self as author* by looking into the impact other strands of identity had on it. It showed that the participant's self-perceived autobiographical identity as a culturally and linguistically minority student influenced how she positioned herself in English writing: There was a tension between what she wanted to position herself and how she actually positioned herself.

Also focusing on the source use and plagiarism-related issue, Abasi et al. (2006) did a multi-case study with five ESL graduate students at a Canadian university focusing on writer identity constructed from their citation behaviors. These five students were divided into two groups: less and more experienced writers. The findings showed that more advanced writers, who had rich academic literacy experiences, consciously represented themselves as intertextually knowledgeable writers in their papers through referencing and citations, while irregular documentation, unattributed borrowed chunks, and unduly “patchwriting” (Howard, 1995) could be found in less experienced students’ papers. The authors argued that writers’ authoritative view of source texts led to the appropriateness of appropriating the source texts. Hence student writers should be aware that source texts should be viewed as “internally persuasive” (Abasi, Akbari, & Graves, 2006, p. 112) rather than authoritative, so that they can have dialogic interaction with source texts in order to better integrate them in their own texts.

Fazel and Shi (2015) investigated six doctoral students’ emerging scholarly identity construction through looking at their citation behaviors strategically and rhetorically. By using qualitative analysis and interviews with the students, the study identified five rhetorical functions of the citations, namely *to claim knowledge, to seek support, to claim importance, to establish a territory,* and *to claim competence,* and three strategies, including *to emulate other writers, to follow professors' suggestions,* and *to mask unfamiliarity with the topic*, which underlie the five rhetorical choices. The study concluded that making academic identity claims in grant proposals for doctoral students who assumed dual professional identities simultaneously – advanced students and future academics, is a process of “becoming”, and “socialization” (Fazel & Shi, 2015, p. 212).

2.5.2.3 Metadiscourse and writer identity

Following the conventions in sociolinguistics, in order to understand writer identity construction, researchers have anchored relevant linguistic features that are found to be helpful, one of them being metadiscourse, which is “central to writers’ representations of themselves and to the organization and presentation of their arguments in their texts” (Starfield, 2004, p. 153), and it is a key feature associated with the success of academic writing (Hyland, 2004b). The framework of metadiscourse can be seen as linguistic resources manifesting the writer’s position towards content and their readers (Hyland, 2004b; Kuteeva, 2011), and metadiscourse constitutes the main approach to understand writer identity in academic discourse (Supasiraprapa & De Costa, 2017). As a crucial element in academic genres, on the one hand, metadiscourse helps writers to guide the readers throughout the writing via the use of various textual links such as transitional markers and signposts. On the other hand, it helps demonstrate the writers’ attitude and stance toward the propositions and content and project the relationship they assume with readers. Metadiscourse is essential in negotiating the relationships with readers, effectively claiming authoritativeness, and achieving communicative purposes when performing rhetorical acts such as arguing, rebutting, persuading, and asserting.

By definition, metadiscourse is “self-reflective linguistic material referring to the evolving text and to the writer and imagined reader of that text” (Hyland & Tse, 2004, p. 2004). Hyland (2018) also points out that metadiscourse refers to a body of linguistic resources writers use to organize text, make statement, position attitude, and engage readers in the text, and he classifies metadiscourse along two main dimensions: interactive and interactional. Interactive metadiscourse entails five sub-categories. Interactional metadiscourse encompasses stance and engagement markers (Hyland, 2005). Each of the metadiscourse markers will be discussed below in more detail. Combining both Hyland's (2005, 2018) frameworks, the metadiscourse model adopted in this dissertation is presented in Table 1.

Interactive metadiscourse markers focus on the internal structure of academic papers and are "interactionally motivated, contributing to the creation and maintenance of shifting interpersonal orientations" (Hyland, 2004b, p. 137). Interactive metadiscourse entails five sub-categories: *transitions* refer to conjunctions used to signal the semantic relations between clauses (e.g., and, but, furthermore, likewise, similarly, so, and therefore), *frame markers* refer to devices that mark the stages and structure of texts (e.g., first, then, to conclude, and my purpose is to), *endophoric markers* refer to other parts of the text (e.g., see Figure 1 or as shown in the introduction section), *evidentials* are mainly about including sources of information outside the text through citations (e.g., XX, 1990), and *code glosses* rephrase or restate information for further elaboration or clarification (e.g., for example, namely, that is to say).

Interactional metadiscourse, on the other hand, signals writers' positioning in relation to the issues discussed in the text and to readers (Hyland, 2005). Interactional metadiscourse contribute to the interpersonal dimension of the text and it includes stance and engagement markers. According to Hyland (2005), stance refers to the textual 'voice' and conveys writers' judgements, attitudes, and commitments, which shows the degree to which writers exhibit their personal authority. It encompasses *hedges,* the employment of which indicates writers' uncertainty to full commitment to a proposition (e.g., perhaps, possibly, may), *boosters,* which, on the contrary, express writers' assertiveness in making a claim (e.g., obviously, clearly, highly), *attitude markers,* which state writers' affective attitude toward propositions, such as surprise, interest, agreement (e.g., unfortunately, significant, interesting, and prefer), and *self-mentions*, words explicitly referring to the writer themselves (e.g., I, my, me, and the author).

Engagement, on the other hand, signals that writers bring readers into the discourse by addressing their expectations, predicting their responses, and guiding them to particular interpretations. Five elements are categorized under engagement markers, including *reader pronouns*, explicitly referring to readers (e.g., you, your, and one), *personal asides*, which interject a statement by inserting their own comment to address readers, usually found after dash or in parentheses, *appeals to shared knowledge*, which writers use to indicate solidarity and mutual understanding with readers (e.g., obviously, common knowledge, we all know that…), *questions* directly involve readers in the knowledge making by prompting readers to think about the questions led by the writer, and *directives,* which are used to instruct readers to perform an action or see things from the perspective prescribed by the writer (e.g., see…, must, should).

Table 1. Analytical framework: metadiscourse features, based on Hyland (2005, 2018).

Metadiscourse	Category	Function	Examples
Interactive	Transitions	Show semantic relation between main clauses	And, however, but, therefore, in addition
	Frame markers	Specify discourse acts, sequences or stages	First, second, in conclusion
	Endophoric markers	Refer to information elsewhere in the text	Noted above, in Section 1
	Evidentials	Cite information from other texts	According to, (Y, 2005)
	Code glosses	Elaborate propositional meanings	Namely, such as, e.g., for example, in other words
Interactional (stance)	Hedges	Withhold commitment and open discursive space	Might, possible, it is suggested, indicate
	Boosters	Express certainty or close dialogue	Definitely, show, demonstrate; certainly
	Self-mentions	Refer explicitly to author/s	I, my, our
	Attitude markers	Express writer's affective stance toward propositions	Surprisingly, intuitive
Interactional (engagement)	Reader pronouns	Refer explicitly to readers	You, one, we, us
	Directives	Instruct readers	See Appendix A, should bear in mind
	Questions	Invite direct collusion from readers	?
	Appeals to shared knowledge	Signal the writer's community membership and solidarity with readers	It is well known that, common knowledge
	Personal asides	Offer meta-comment	By the way, personal comment appearing in parentheses, dashes, or commas

There has been proliferative research conducted on metadiscourse, and stance and engagement in various genres, such as research articles (Hyland, 2005, 2018; Hyland & Jiang, 2016, 2018, 2022; Hyland & Tse, 2004), graduate theses and dissertations (Lee & Casal, 2014;

Thompson, 2012), academic blogs (Hyland & Zou, 2020; Zou & Hyland, 2019, 2020, 2022), to name only a few. Investigating writer identity through the lenses of metadiscourse is relatively fewer in current writing scholarship. There are a few that addressed writer identity through metadiscourse. For example, Burke (2010) examined ways of how six Korean students constructed writer identities ideationally, interpersonally, and textually using metadiscourse framework. According to the research, the Korean students highlighted two rhetorical functions academic writing is expected to achieve: a) present knowledge with the use of numerous citations and b) guide readers into their ideas with many transitions in their papers. In order to accomplish these functions, the study found that textual metadiscoursal markers such as transitions, code glosses, and evidentials were frequently used in these students' papers, while interpersonal metadiscoursal markers such as hedges and boosters were used less frequently, which made their writing fail to construct these students as authoritative academic writers.

In another dissertation research, Taj (2017) conducted a case study and studied the identity construction in four Arabic students' academic writing who was working on their doctoral degree in the UK. The textual data in the study included various sections of research writing, but the author found not all the participants used the identity markers (i.e., metadiscourse markers) consistently throughout their writing. Reader reference was one metadiscourse item that was under investigation that all participants avoided to use. In another case study, Supasiraprapa and De Costa (2017) analyzed identity building in two second language MA students' teaching philosophy statements drawing on metadiscourse framework. Their study revealed that in order to build identities as a graduate student and teacher, the two students employed almost all the metadiscourse markers in their teaching philosophy statements.

Further, using metadiscourse as the analytical framework, Castillo-Hajan et al. (2019) drew on 50 persuasive essays written by ESL senior high school students and found that self-mention was found in all samples of the data, which served as a metadiscoursal feature securing the writer's identity. In contrast, the study found that attitude markers, i.e., *especially*, *more importantly*, and *unfortunately*, were the least frequently used metadiscoursal marker in their corpus. Tessuto (2008) compared writer identity in 30 Introduction sections in law research articles written by 20 native writers and 10 non-native writers. Focusing on the interactional metadiscoursal features, the results showed writer identity is achieved through explicit use of self-mentions and the authors' strong inclination to self-citation.

2.6 Doctoral students' identity

Identity in recent years has become one of the major foci of research on doctoral studies, and identity development has been considered as the key outcome of doctoral studies (Inouye & McAlpine, 2019). Education at the doctoral level "is the socialization of individuals into the cognitive and affective dimensions of social roles related to the practice of learned occupations" (Weidman & Stein, 2003, p. 642). For doctoral students who are amidst an uncertain and transitioning identity space (Fazel & Shi, 2015), namely being advanced student writers and future scholars simultaneously, how to project identities responding to particular rhetorical situations is exceedingly important. For doctoral students, to construct an academic identity is about "being able to think, act and develop their own research voice even if they are still considered as students both by the institution and by themselves" (Castelló et al., 2013, p. 445). As has been demonstrated by a few previous studies, constructing a scholarly identity in text is extremely challenging for doctoral writers, because representing oneself this way is essentially a reflection of one's social identity as an academic author and researcher (Ivanic, 1998). Green (2005) points out that identity formation is equally important as knowledge production in doctoral enculturation. Scollon also suggested that academic writing "is as much the construction of an authorial self as the presentation of fact" (1994, p. 34).

Writing serves as one of the most important markers of the measurement of doctoral students' achievement and graduate competency (Aitchison et al., 2012), and it is through writing that doctoral writers' identities are forged and they become visible to the disciplinary and academic community. Claiming authority through writing requires doctoral students not only to have a comprehensive grasp of field knowledge but also to be able to critically amalgamate the knowledge, make knowledge contribution, and maintain an authorial identity. Text and one's identity are inextricably linked and everything one writes convey to readers the persona or impression of the writer (Ivanic, 1998). As Kamler and Thompson (2014) states, "The text is an extension of the scholar, a putting of 'self' out there which is either successful – or not" (p. 15). To produce academic texts that not only demonstrate writers' knowledge but also make contributions to the field is an essential goal for doctoral writing (Casanave, 2019), which makes constructing a credible and scholarly image in texts crucial for doctoral writers.

It has been argued that though L2 writers are not linguistically disadvantaged compared to Anglophone writers (Hyland, 2016), writing academic papers not in their first language is still challenging and contested to L2 writers (Botelho de Magalhães et al., 2019; Langum & Sullivan, 2017). Though doctoral students are considered writers of advanced level, writing still constitutes a major pressure for them (Maher et al., 2008) and demonstrating scholarly voice is not without difficulty. Aitchison et al.'s (2012) study shows the struggles, adjustments as well as the affective challenges (e.g., frustrations and uncertainties) doctoral students experience as they try to master the advanced linguistic practices in academic writing and positions themselves as independent scholars in their discourse communities.

To some L2 PhD students, writer identity is even a notion they are not aware of. For example, Taj (2017) found four Saudi students working on their PhD in the UK sometimes were unconscious of the concept of writer identity, as some discoursal decisions could not reflect their reflective choices. The study also found that these students tended to over rely on advice and guidance from their supervisors, as is influenced by the Saudi ideologies which attached importance to the role of teacher in the learning process. Also, writers in the study were found to recourse to translations from Arabic to English to compensate for their poor language skills, which would unavoidably influence the construction of writer identity, as their English texts were mediated by Arabic patterns of thought and style.

Zhang and Hyland (2021) point out the importance of apprenticeship in the shaping and construction of scholarly identities for doctoral students. That is to say, the guidance and feedback from "community-sanctioned experts" (p. 1) can scaffold doctoral writers' construction of identity as a competent, credible and knowledgeable community insider through texts (Aitchison et al., 2012; Inouye & McAlpine, 2019). Evidence has also shown that explicit instruction on authorial stance-taking can raise L2 writers' awareness and improve their writing. For example, Chang and Schleppegrell (2016) aimed to assist seven ESL doctoral students to develop more effective authorial voice by making explicit linguistic resources for claiming authorial stance. Results showed that these students made improvement in terms of rhetorical move structure and stance deployment after they were instructed explicitly using a concordance tool. Boz (2016) also argued that academic writer identity was formulated under the influence of teacher feedback and grading practices. In the study, the departmental teachers had access to essential discipline-specific information that could guide students regarding constructing academic identity, but the study pointed out that most of the students in the study did not have adequate contact with course lectures to constitute apprenticeship in order for them to seek oral or written feedback for previous assignments before turning in their next one.

Anderson (2017) conducted a multi-case study over 16 months which involved seven Chinese doctoral students at different stages of their programs at a Canadian university. Using semi-structured interviews, student-generated narratives, and some written feedback on their academic writing, the study outlined internal and external factors impacting these students' socialization into the target discourse/academic communities. Externally, for example, the degree to which the supervisor was helpful and supportive made a considerable difference on students' perception regarding their academic identity; however, how doctoral students internally positioned themselves in relation to their peers could interfere with their interpretation of themselves, which, in most cases, brought pressure and frustration.

For research writers, it is essential to display their domain knowledge and situate their voice within the broader disciplinary and academic community. Projecting an authorial identity through writing is a key component of academic success. To this end, writers need to balance the considerations about their personal values (Zhang & Hyland, 2021) in relation to "dominant ideologies, norms, and social expectations" (Canagarajah & Matsumodo, 2016, p. 3). This apparently requires writers to be able to make use of linguistic resources to fulfill the expectations in terms of both knowledge presentation and knowledge production in their research writing. Doctoral writing poses considerable challenges to PhD students, which has been the focus of a number of previous research. This positions "doctoral writing and identity as worthy areas of inquiry with important practical implications for pedagogy" (Inouye & McAlpine, 2019, p. 4).

Deng (2012) explored the Discussion and Conclusion sections written by six Chinese social science doctoral students at different stages (from draft stage to revision stage) who were studying in Hong Kong for their PhD degree. Based on the interview data with the students, the study found the students' identity exhibited a developmental pattern from being a novice student writer, to a more skilled academic writer, and finally to claim full membership of the community. The result also revealed that the different amount of social interactions with the target discourse community at different phases of their PhD process affected the students' acquisition of disciplinary genre knowledge, which in turn impacted their identity construction.

Though writer identity research regarding L2 doctoral writers is not large in number, the extant body of research has fundamentally accentuated the importance of sociocultural factors in shaping writer identity (*discoursal self* and *self as author*). L2 users' linguistic, social, cultural, gender, racial, and ethnic identities (*autobiographical self*) mediate their access to linguistic resources and interactional opportunities available in their L2 (Pavlenko, 2000, 2001) that help them build identity in writing. Combining the insights gained from writer identity research in other cultural and rhetorical situations, this study aims to look into writer identity construction among ESL doctoral writers' papers in the US settings, with the goal of disclosing the linguistic patterns in the writers' text that contribute to the construction of writer identity, and the influence of multifaceted factors, including contextual and personal factors on the construction of writer identity.

2.7 Research gaps

In the past several decades, identity research in the field of applied linguistics has witnessed some development. Theoretically, it has been approached from different philosophical backgrounds such as social constructionism, social constructivism, poststructuralism, sociolinguistics, sociocultural perspectives, and so on. The seminal conceptual framework from Ivanič (1998) has laid a solid foundation for subsequent writer identity research. This framework has also been conceptualized from new dimensions, such as different aspects of writer identity corresponding to timescales (e.g., Burgess, 2012; Burgess & Ivanič, 2010). Writer identity is co-constructed by writer and readers and mediated by semiotic signs (Burgess & Ivanič, 2010; Matsuda, 2015). What's more, the analytical frameworks drawn to analyze writer identity textually provided an important lens and guidance for researchers to locate this intangible construct in tangible discourse. Methodologically, corpus tools, text analysis, qualitative analysis, case studies, and interviews have been employed to provide understandings of L2 writer identity from various perspectives. These contributions have advanced writer identity scholarship and identity research. However, there are some aspects that have been relatively underexplored. In the remainder of this section, I elaborate on these research gaps and how previous studies have informed the current dissertation.

Based on prior research, projecting an authorial identity to fulfill certain rhetorical effect is a challenging skill that is beyond the mastering of grammar and vocabulary for L2 writers, and the challenge is also caused by contextual factors. There has been limited research examining how L2 doctoral students construct their identities in the ESL social and academic environment. Therefore, we have limited understandings toward how they represent themselves in writing and why they do that that way. From a social constructivist view, writer identity is constructed through socially shared repertoire for meaning making (Matsuda, 2015); yet the linguistic resources are appropriated by individual writers selectively to respond to particular rhetorical situations. However, how L2 doctoral writers mediate, manipulate, and balance the resources during their writing process is a question that needs to be investigated. Supposedly, there might be tensions, compromises, objection, resistance, or conformity between the identity constructed and all the factors. But how these negotiations manifest themselves in L2 doctoral writers' academic writing and why this group of writers build identity in written discourse the way they do? In light of this, in order to shed light on the L2 doctoral students' writer identity practices, this study aims to examine both the textual features constructing the writer identities, and the reasons behind their choice of manifesting themselves in texts the way they do.

Further, as identity is a highly context-sensitive construct, how doctoral writers from different disciplinary backgrounds construct their identities similarly or differently has yet to receive attention. How linguistic choices help construct their identities through the texts and what factors would influence their construction of identities are not clear yet. Given this, this dissertation aims to explore L2 doctoral students' identity construction from two contrastive disciplines: hard and soft disciplines, and attempts to address this research gap in regard to how disciplinary background (contextual factors such as discourse communities and genres) influences identity construction. Moreover, how agency plays a role in the construction of identities seems not a concern in the current body of literature on L2 doctoral students' identity construction. This dissertation also intends to explore how doctoral writers exert their agency to construct identities.

Previous research has also investigated writer identity through multiple research methods. In order to reflect the complexity of this construct, combining textual analysis and interviews, which can generate the writer's ideas regarding how they present themselves in their writing and why it is the way it is, can often purvey researchers with valuable insights. Interview is a common way of eliciting participants' ideas toward an issue in question. While previous research has considered the impressions brought by a text to readers and readers' perceptions of the writers' real-world identity through reading, i.e., reader-based approach (e.g., Matsuda & Tardy, 2007; Tardy & Matsuda, 2009), writer-based approach has not been used to explore writers' sense of the papers they compose. That is, from the writer's perspective, how they perceive the papers they write, and why they construct themselves the way it is has barely been investigated yet. After all, writers are the agents that produce texts: They are the ones who are most familiar with how they write a text, what factors shape their writing, and how their agency is played out and affected during the composing process. Although readers are those who ultimately perceive the construction of writer identity (Matsuda, 2015), we still need to understand writers' perceptions, which may or may not be congruent with readers' perceptions. These perspectives are of great importance in demystifying writer identity construction and are indispensable in helping provide a comprehensive understanding regarding writer identity, and how and why it is formulated. Given its importance, this study will adopt writer-based approach that attempts to uncover writer identity from writers' perspective. Additionally, with the temporal dimension being an indispensable constituent in identity construction longitudinal studies tracking the developmental trajectories are also a research method that will be adopted in the current research.

In order to understand how L2 doctoral writers construct themselves through writing (their *discoursal self* and *self as author*), how the individual, and contextual factors influence identity construction, this dissertation aims to investigate the following questions:

2 How do L2 doctoral students construct writer identity in their papers of different genres?

1a: What metadiscoursal features does their writing have that contribute to the construction of writer identity?

1b: What are the patterns of metadiscourse use in identity construction in the doctoral students' papers across different written genres?

3 How do contextual factors and agency shape the L2 doctoral students' construction of identity in their papers?

CHAPTER 3: METHODOLOGY

3.1 Rationales for the research design

The major objectives of this dissertation are to investigate how writer identities are constructed through metadiscourse in L2 doctoral students' academic writing of different genres and across disciplines, and what factors, both intrinsically and extrinsically, influence these students' identity construction in writing. In order to address these "how" and "why" questions, this dissertation adopted a qualitative multi-case research design. Qualitative research has been characterized by the following common features: multiple data sources, small-scale design, non-generalizability, an inductive analysis of data, an interpretative methodology, and a reliance on researcher's subjectivity and reflectivity (Bogdan & Biklen, 1992; Eisner, 1991; Merriam, 1988, 2002), and "emic (qualitative) research is much more heterogeneous in its ontological and epistemological assumptions" (Markee, 2013, p. 3). As Denzin and Lincoln (2005) stated:

> Qualitative research involves the studied use and collection of a variety of empirical materials — case study; personal experience; introspection; life story; interviews; artifacts; cultural texts and productions; observational, historical, and visual texts—that describe routine and problematic moments and meanings in individuals' lives. Accordingly, qualitative researchers deploy a wide range of interconnected interpretive practices, hoping always to get a better understanding of the subject matter at hand. (pp. 3-4)

More specifically, the rationales for employing a qualitative multi-case study are as follows: First, this research highlights the role of context, as identity is understood as a contextualized, fluid, and heterogenous construct; understanding it within a bounded context will provide us with insights and complexities. Qualitative research aims to seek to gain an understanding of a phenomenon while taking into account potentially relevant influences. It emphasizes "understanding the particular context within which the participants act, and the influence that this context has on their actions" (Maxwell, 1996, p. 17). Second, as writer identity construction is a largely idiosyncratic practice (Ivanic, 1998), this study is interested to investigating the individuals' traits in terms of their construction of identities, and what factors contribute to their identity constructed in the papers. To address these goals, a qualitative design would be helpful in disentangle these issues.

Case study is designed to be particularly suitable for "situations where it is impossible to separate the phenomenon's variables from their context" (Merriam, 1988, p. 10), and it aims to uncover insight, discovery, and interpretation rather than test hypothesis. Case studies can be characterized by the following features (Merriam, 1988): *Particularistic* means case studies explore particular situation, event, and phenomenon. *Descriptive* means case studies present a "rich" description of the phenomenon under investigation. *Heuristic* means case studies offer insights and illuminance and enlighten readers in regard to the phenomenon. *Inductive* refers to that case studies are based on inductive reasoning, with the hypothesis and concepts emerging from the data, which in turn are situated in context. A multiple-case method is used to reveal the complexities and make possible comparisons between different participants, which is suitable for this dissertation study. First, the essential nature of qualitative research and case study research allows the researcher to study writer identity in a more contextualized manner and to gain holistic understandings from individual writers and their immediate contexts. Second, when collecting data, multiple sources of data were collected in this dissertation study in a naturalistic setting, without interference with the research environment. Third, writer identity usually develops along an extended period of time; therefore, a longitudinal study design was employed in order to address this developmental trajectory. Lastly, a multi-case study design is considered because factors such as disciplines and one's prior life history can account for different patterns in identity construction unique to individual writers.

The sections below concentrate on the methodology I adopted for the study from the following aspects: introduction of the research context, rationales for the selection of the participants, procedures of conducting interviews and collecting textual data, and steps of data analysis. Concerns related to the validity of qualitative research design are addressed toward the end of this chapter.

3.2 Context of the study

From social constructivist perspectives, writing practices cannot be fully understood without situating them in particular sociocultural and institutional contexts. Given its importance, in this section, I first describe the context of the study, followed by the presentation of the selection of the participants, researcher subjectivity, and ethical issues.

3.2.1 Research setting

This study was designed to take place at a public research university in southeastern United States. The university offers more than 250 majors at the master's, specialist, and doctoral levels, and has about 5,000 international students, making up about 8.6% of the student body (College Factual*,* 2022). International students come from at least 50 countries in the world, with the largest groups being India, China, and Saudi Arabia (College Factual*,* 2022).

The two programs the three participants came from were both doctoral programs at this university, one being Transportation Engineering and the other being Applied Linguistics. The two students from transportation engineering were supervised by the same advisor, who required his students to publish at least three journal articles to fulfill one of the graduation requirements. The program accepted applicants with or without a master's degree; one of the two participants, therefore, started her PhD study directly after obtaining bachelor's degree. The linguistics program did not stipulate that its doctoral students should publish in order to graduate, but it does only accept applicants who have a master's degree in related areas. Students in both programs completed courses in the first two years of doctoral study, while the remaining years were devoted to dissertation research.

3.2.2 Selection of participants

The rationales for participant selection are as follows: First, the selection of cases was based on purposeful sampling (Duff, 2008). From an accessible population within my social network, I selected participants that fit the purposes of this dissertation study. Second, as writers in different disciplines represent themselves in different ways (Halliday & Martins, 2003), in order to address the ways disciplinary norms and conventions contribute to self-representation in doctoral writers' academic writing, students from two disciplines were recruited: transportation engineering and applied linguistics, which could represent the hard disciplines and soft disciplines respectively (Biglan, 1973; Sealey & Carter, 2004). Third, in order to examine how different stages of doctoral study influence identity construction, I recruited participants from various stages of their PhD journey: My participants were in the first year and fifth year in their PhD programs.

In Spring 2021, I sought approval from the Institutional Review Board (IRB) at the institution of the research site as my study involves human as participants. It was determined that my proposed study met the criteria for exemption from IRB review. What needs to be pointed out is that the IRB staff approved that this type of research only required oral consent; therefore, I sought consent from my participants verbally. The exempt letter can be found in Appendix C.

Actually, seven participants initially agreed to participate in my research, but four of them dropped out due to various kinds of reasons. Attrition is a major issue in qualitative longitudinal research (Ahern & Le Brocque, 2005) due to the fact that the data collection process lasts for an extended period of time. After the seven participants agreed to participate in my study, I conducted background interviews with six of them in Fall 2021. In the subsequent semester, I collected a few papers from them, and did paper interviews for one or two rounds. As I told my participants I needed their papers written for three different purposes, some did not have sufficient papers. Yet, some participants opted out in the middle of the data collection because they felt uncomfortable sharing their writing data. Some participants relocated to another State in the U.S., which made data collection less accessible. Since participation in this study was completely voluntary, only three participants shared with me their papers of different genres and completed the corresponding interviews. Therefore, in this dissertation, I report results related to these three participants.

Among the three doctoral students participating in the study, two students from engineering and one from applied linguistics. The two participants from the engineering department were recruited through a friend of mine, who knew of my dissertation study and introduced her interested colleagues and friends to me. The one from applied linguistics was recruited from my own social network. The participants from engineering came from different levels of study: one just started their doctoral degree, the other was a senior PhD student who had passed the Qualifying Exam and was in the dissertation writing stage. The applied linguistics student was in the first semester of his doctoral study. All the participants were originally from China. Below I present the profiles of the three participants which were generated from the background interviews I conducted with them (as will be reported in detail later in this chapter), starting from the two from transportation engineering. The profiles of each participant are displayed in Table 2.

3.2.2.1 Momo

Momo was 25 years old at the time of data collection. When data collection began, she was a first year PhD student, while she just started her second year at the later stage of data collection. She came from China; her native language is Chinese. Momo completed both her bachelor's degree and master's degree in her native country. Her major was transportation engineering in both stages of study. Momo's undergraduate program was a bilingual program where the courses were delivered in Chinese and English, and the exams in each semester were written in English and students were required to answer in English. Their assignments, in a similar vein, were also required to be done in English. Due to this exposure to English in and after class, Momo's English proficiency was relatively higher than that of her undergraduate peers who were not in the bilingual program. After graduating from the undergraduate program, Momo chose the same major for her master's degree at a key university in China. Right after obtaining her master's degree, in Fall 2021, Momo started her PhD degree in the civil engineering program at the research cite. Her current goals after obtaining the PhD degree were to stay in academia and to be a researcher, which, according to her, might require her to work as a postdoctoral fellow first after graduation.

Momo's TOEFL score was 98 out of 120, with writing being 24 out of 30. She took an academic writing course in the undergraduate program, which mainly focused on citation format, avoidance of plagiarism, correspondence writing, and so on. Her master program did not offer academic writing related course, but she learned relevant writing knowledge from the TOEFL and GRE courses she took off campus. As to the academic writing experience, Momo wrote her undergraduate graduation paper in English, as required by the bilingual program. Her master's thesis, however, was written in Chinese. During the three years in her master's program, she worked collaboratively with her mentors and senior graduate students to conduct research, and write and publish papers. Some of the papers she wrote were in Chinese and the others in English. She learned knowledge about academic writing from performing it: her mentor had helped her exceedingly in terms of logic, flow and wording in paper writing, and she improved greatly through this collaboration process.

In the doctoral program at the research site, Momo first was taking courses mandated by the program. However, she did not write extensively in those courses where the assignments primarily centered on algorithm, data processing, and model calculation, with scarce attention given to writing itself. Hence the academic writing she had been working on was mainly for the purpose of publication. She attached great importance to publications as publishing three articles was a mandatory rule in her program; meanwhile, publications would contribute to the advancement of her career in the future. Normally, she used tools such as translation applications, paraphrasing applications, and Grammarly to enhance precision and accuracy in wording and grammar. But she admitted that she had most problems with the development of ideas, which was also reflected from her mentor's feedback on her papers. In terms of other research experiences, Momo was an anonymous reviewer offering comments on manuscripts for journals. The papers collected from Momo for this dissertation study were for the purposes of publication and conference presentation, and she aimed to eventually develop the latter into a publication.

3.2.2.2 Nancy

Nancy was 25 years old at the time of data collection. She came from China and her native language is Chinese. She acquired her bachelor's degree in China, and right after that, she came to the U.S. to pursue her doctoral degree in transportation engineering. What needs to be mentioned is that Nancy came from the same undergraduate program as Momo did – a bilingual program in transportation engineering in China. Upon completion of her bachelor study, she did not choose to study for a master's degree first but to "face the challenges" and chose the doctoral path. Writing in English in her undergraduate program mainly centered on discipline-related genres, such as course assignments, report, etc. Since she planned to study abroad in her undergraduate study, she studied for IELTS and GRE off campus, so according to Nancy, her writing proficiency was "above average" compared to her peers. Nancy's IELTS score was 7 out of 9, and the Writing section was 6 out of 9. Since Nancy did well in the tests, she also worked part-time as a GRE teacher for Quantitative reasoning in China.

However, since she started the doctoral degree after graduating from college, where she did not receive adequate research training, she felt the initial stage of PhD studies was highly challenging. For example, in terms of reading papers, she found it difficult in terms of not only language but also the knowledge and content. She had two advisors for her doctoral study: one was from the transportation engineering program, the other from a research center she worked in as a research assistant, which could be considered as two different communities of learning (Castelló et al., 2013). As a graduate research assistant working in the research center, Nancy collaborated with her advisor on research projects and wrote up report mainly to fulfill the job funded by certain government departments. Nancy wrote proposals and reports extensively when working at the research center, and her advisor offered feedback on her writing that helped her improve significantly. Meanwhile, Nancy's advisor from the research center also hired a native speaker of English to proofread the proposals and reports, who primarily focused on flow of ideas, grammar, and wording. Given that she received funding from the research center, she worked the majority of the time of her PhD journey with the advisor instead of the program advisor, whom Nancy only met regularly at the dissertation writing stage.

It is worth mentioning that in the third year of her PhD, Nancy took a course called *Research Methods*, which she found very enlightening and helpful. This course covered how to write proposals and dissertation, how to read and write journal papers, and how to do presentations. At the beginning of data collection (Fall 2021), Nancy had published two articles. However, she stated that working at the research center influenced her choice of future career path: she was not interested in staying in academia, and planned to look for some practical job in the industry in the US after graduation. The papers collected from Nancy included one article written for publication, one research report she wrote working at the research center, and her dissertation.

3.2.2.3 Adam

Adam was 24 years old when the data collection began. He came from China and his native language is Chinese. He completed his bachelor's degree in TESOL at a Chinese-foreign cooperatively founded university. Being cultivated with an interest in linguistics from the courses he took in the undergraduate program (e.g., *Language and Society*, *Language and Culture*), he pursued a master's degree in applied linguistics in the UK after graduating from college. Motivated by a future goal to work at the tertiary level, which generally required a doctoral degree, he came to the US to further his study of PhD degree in applied linguistics. During his undergraduate study, he took an academic writing course named *Advanced English Composition* where students were taught basic academic writing paradigms, such as how to write a bibliography. He also took some linguistics related courses such as *Corpus Linguistics* and *Sociolinguistics*. The courses offered in the undergraduate program were delivered in English, and the assignments he completed were also in English. The two stages of study both prescribed Adam to write a research paper to fulfill one of the graduation requirements. Therefore, since undergraduate study, Adam had had experience related to conducting empirical research and academic writing.

When data collection began, Adam was in his first semester of doctoral studies, and he needed to take three courses that semester. Assignments from each course encompassed different genres, from personal reflections, book reviews, to research reports. Writing samples collected from Adam included one book review, two research reports written for two courses. The two courses were taught by the same instructor, with whom Adam shared common research interests. And in the remaining years of his PhD, Adam aimed to collaborate more with the instructor.

Table 2. Profiles of the participants

Name	Gender/Age	Program	Year of doctoral study	Native language
Momo	Female/25	Transportation engineering	1st – 2nd	Chinese
Nancy	Female/25	Transportation engineering	5th	Chinese
Adam	Male/24	Applied linguistics	1st	Chinese

Participating in research that makes individual writers reflect on linguistic and writing issues is beneficial in essence (Duff, 2008), and it is suggested that letting learners talk about their perceptions can facilitate their consciousness regarding the construct under investigation (i.e., writer identity) (Ivanič, 1998). For example, in the interviews, Momo realized that she did not understand the connotations of such verbs as “demonstrate”, “show”, “indicate” and “suggest”. Our interviews and discussion helped her gain the awareness of academic word usage, so she could select the words she originally thought were same in meaning in a more skillful way in her future writing. Also, participants were compensated for their time. When interviews were conducted face to face, I brought a cup of coffee for my participants. When data collection was complete, I sent each participant a $50 Amazon gift card.

3.2.3 Researcher subjectivity and reflexivity

Reflexivity is usually “associated with a critical reflection on the practice and process of research and the role of the researcher… It acknowledges the mutual relationships between the researcher and who and what is studied” (Lichtman, 2012, p. 165). Common to other social practices, research is value-laden. Human beings cannot set their subjectivity and values aside (Johnson, 1992, p. 32) when performing any social acts, and doing research is no exception.

Hence the researcher's subjectivity is meant to come into play in this qualitative multi-case study during all phases including selection of topic, data collection, data analysis, and the final write-up of the report. After all, whether objectivity exists or not is a question that remains debatable in different theoretical orientations (Duff, 2008). Lichtman (2012) points out that in qualitative research, it is impossible to find objectivity, and researchers should not strive to be objective but "face head on the subjective nature of their role" (p. 159). Considering the unavoidable involvement of subjectivity, qualitative research uses the researchers' subjectivity – including their tacit knowledge, and insights – as an instrument (Angélil-Carter, 1997; Ely, 1991). Data are "mediated through this human instrument" (Merriam, 1988, p. 19). It can make full use of the researcher's subjectivity to "'dialectically and dynamically' balance letting the object speak, and avoiding distortion" (Angélil-Carter, 1995, p. 37).

Specifically, that I chose to focus on doctoral writers' identity in academic written discourse was mainly motivated by my personal experience as an international PhD student who had to transition from a sociocultural and academic environment in China to the US after I came to the US in 2018 to pursue my doctoral degree. In my first year of PhD, I found there were all kinds of challenges in my brand-new social identity – being an L2 doctoral student. I was overwhelmed, frustrated, stressed and at a loss. Somewhat incidentally, when I was preparing for one of the course papers the first semester, I saw Botelho de Magalhães et al.'s (2019) article in the *Journal of Second Language Writing*, which was about two L2 doctoral students' experience of being socialized into their target discourse community. Motivated by my experience, primarily for the sake of knowing more about myself and my identity, and the academic socialization process, at that time, identity became a topic that greatly intrigued me. Since I have been interested in writing research, writer identity as a research topic has captivated my attention.

Over the past five years, I have grown and made some progress as a researcher and I have received some feedback regarding my academic writing. That made me realize that my writer identity has developed over the past a few years. My own experience cultivated my interest in getting to know what other L2 doctoral wrote could demonstrate their identities. I believe my experience of being socialized into an English-medium academic context is not uncommon among all the L2 students working on a PhD degree in the US. Therefore, I chose to focus on L2 doctoral writer's identity as my dissertation topic. Further, the three participants in this dissertation research all came from China; they were in the US to pursue their doctoral degree. The same linguistic and cultural background and similar academic experience about switching from one academic environment to another helped me build rapport with my participants. I assumed the role of *researcher-as-befriender* during the data collection process (Candlin & Sarangi, 2003). Since we shared a lot in common, they were not just my participants from whom I could mine data. More importantly, I treated them as friends, which could help build a relaxing and encouraging atmosphere to let them talk freely about their ideas on the papers.

3.2.4 Ethical issues

Ethical issues can be a main concern when conducting case studies, as they will unavoidably involve thick descriptions of each case, which can reveal one's core information that might be identifiable. According to Merriam (1988), ethical issues can emerge in qualitative case study research at two points: during data collection and results dissemination. In order to protect participants' security and privacy, first of all, I submitted my proposal to the Institutional Review Board's (IRB) for review from the university where I am at before carrying out the study, and the protocol was determined exempted for review. The letter for exempt determination can be found in Appendix C. Second, pseudonyms were assigned to the participants to protect their privacy, and the participants were given the options to choose a pseudonym for themselves if they did not prefer the name I created for them. Further, the titles of their papers were significantly shortened in order for their real identities not to be revealed by any chance. Third, upon successfully getting approved by the IRB, I ensured that my participants were fully aware of the purposes of the study, how long the study would last, time intervals between our meetings, and what information they would need to reveal. Above all, participating in the study was totally voluntary. If there were any circumstances that prohibited them from continuing, they could opt out without any consequences on their part. These steps were taken in order to make sure harm was minimize and benefits were maximized to my participants (De Costa et al., 2021). All participants were supposed to be protected from harm and embarrassment (Denzin & Lincoln, 2005).

With regard to the dissemination of the data and the findings, the participants were informed that the texts collected from them and interviews recorded would be used for research purposes only and would be kept confidential. They were also notified that I was the only person who had access to their data, including their papers and our interview recordings. After I finished writing the analysis on each case, the interpretations related to the participants were sent back to them for reading and checking; if they objected to the revelation of certain information, it was revised and/or removed in order to protect their "right to privacy" (Cassell, 1978, p. 141). The results were only disclosed if my participants agreed I did not misinterpret their self-reported information.

3.3 Data collection

As noted by Merriam (1988), "case study does not claim any particular methods for data collection or data analysis. Any and all methods of gathering data from testing to interviewing can be used in a case study, although certain techniques are used more often than others" (p. 10), while on the other hand, the researcher will not "manipulate variables or administer a treatment" (p.17). While embracing various kinds of naturalistic data with the progression of the implementation of the study, in this dissertation, materials collected from the three doctoral students' papers and interviews are considered the primary sources of data, complemented by other supplementary materials (e.g., course syllabi, student's research log, advisors' feedback, journal reviewers' comments, and so on) to provide additional information.

3.3.1 Textual data

Textual data in this study encompassed students' papers of different kinds they wrote during the doctoral studies, and corresponding supplementary materials such as feedback on papers and/or guidelines. The reasons for this selection were as follows: First of all, doctoral students usually write papers for different purposes, such as for course requirements and publications. Collecting different types of writing aimed to address the questions which seek to elucidate what patterns their self-representation (*discoursal self* and *self as author*) demonstrated under the influence of genres and targeted audience (e.g., *possibilities for selfhood*). It would be interesting to find out the similarities and differences in regard to identity construction in academic writing for different purposes written by individual doctoral students.

Textual data mainly entailed the three doctoral students' writing of different kinds. Next, I introduce data collected from each participant one by one. In order to protect the participants' identity and privacy, a brief name indicating the topic of their paper was given to each of the paper (since some works had got published, the information about participants may be disclosed if exact titles are provided). Further, it is worth noting that papers collected from the engineering program were mostly co-authored. It is often argued that identity constructed is an individualized practice that tends to take on idiosyncratic characteristics (Ivanic, 1998). This dissertation, however, selected co-authored writing sample from the participants due to the following considerations: First, the co-authored papers collected were written in an authentic context, fulfilling the writer's academic goals, such as for publication and/or job requirement. This was in line with the rationales for data collection following the theoretical perspective of social constructivism, which highlight that writers' identity, as a construct, is shaped in sociocultural and academic context. Second, team publishing is generally a norm in hard disciplines as the researchers are more likely to work colloboratively on a projects (Laband & Tollison, 2000). Third, this dissertation chose the academic papers (if they had more than one author) that were drafted and revised by the participants (i.e., the participants were the first author), with additional resources, such as advisors and seniors offering comments and feedback to polish writing. This way, I could focus on how the participants manipulated resources to construct their identities under the influence of multiple factors. Considering that academic writing, more often than not requires multiple rounds of revisions, which are influenced by heterogenous factors, collecting papers with co-authorship would offer more insights in regard to how identity is formed under particular socio-cultural and academic environment.

The textual data collected from the participants are displayed in Table 3.

Table 3. Information about three participants' textual data

Participants	Types of papers	Supplementary materials
Momo	1. Manuscript for publication – Paper about roundabout transportation (P1)	Drafts of the manuscripts
	2. Conference paper - Paper about a survey (P2)	Research log; Conference reviewers' comments
	3. Manuscript for publication - Paper about a survey (P3)	Research log; Co-authors' comments and feedback
Nancy	1. Manuscript for publication – Paper about lighting at night (P1)	Reviewers' comments
	2. Report for government (P2)	Feedback from supervisor and proofreader
	3. Dissertation (P3)	Dissertation template
Adam	1. Course assignment - Book review (P1)	Syllabus; a comparison paper on the book reviews Adam consulted
	2. Course paper - Paper about university websites (P2)	Syllabus
	3. Course paper - Paper about a social media platform (P3)	Syllabus; instructor's feedback

3.3.1.1 Momo

Momo sent me three papers in total, two for the purpose of publication and the other was a submission to an annual conference in her field. One of the papers for publication was about automatous transportation at a roundabout (P1); the conference paper was a survey study (P2), which then was developed into a paper she aimed for publication (P3). She wrote the papers at different times: The one about roundabout transportation (P1) was written in the first semester of her PhD, while the survey paper for conference presentation (P2) was written at the end of first year; the survey paper for publication (P3) were written at the beginning of her second year. It needs to be mentioned that the roundabout paper (P1) was Momo's collaboration with her advisor from her master's program in China. By the end of data collection, P1 was published. All three papers were co-authored, with Momo being the first author doing the drafting, model trialing, and revising, and the co-authors, including the seniors in her program and her advisor, offering comments and feedback on her papers and helping her revise. Meanwhile, I also collected journal reviewers' comments on the roundabout paper. For the survey study, Momo's research log, her co-authors' comments and feedback, conference reviewers' comments were collected. More specific information about Momo's papers is presented in Table 4.

Table 4. Description of Momo's papers

Types of papers	Word count	Dates completed	Descriptions
1. Manuscript for publication – Paper about roundabout transportation (P1)	6,050	Fall 2021	Co-authored; first author
2. Conference paper - Paper about a survey (P2)	5,722	Summer 2022	Co-authored; first author
3. Manuscript for publication - Paper about a survey (P3)	4. 6,845	5. Fall 2022	6. Co-authored; first author

3.3.1.2 Nancy

I collected three papers from Nancy, all for different purposes. One was an article submitted to a journal (P1), which eventually got published in 2019; the second one was a research report (P2) she and the research team worked on for the state government around July

2021. She was the one that was responsible mainly for the drafting and revision of the two pieces, with her co-authors or supervisors providing feedback. She was the first author of the journal manuscript (P1), whereas for the government report (P2), she was the third author, with her senior co-workers at the research center being the first two authors. For this P2, according to Nancy, she was the one that worked on the writing up of the final report, whereas her co-workers mainly offered comments helping her revise. The third piece was her dissertation (P3), which she defended in Spring 2022. It needs to be pointed out that Nancy's P3 consisted of three independent studies, with the first two published by the time of data collection. The first study in her dissertation was a revised version of her P1. Additional data from Nancy included journal reviewers' comments, feedback on report from proofreaders and her supervisor, and the dissertation template. More detailed information about Nancy's papers is presented in Table 5.

Table 5. Description of Nancy's papers

Types of papers	Word count	Dates completed	Descriptions
1. Manuscript for publication – Paper about lighting at night (P1)	6,184	2019	Co-authored; first author
2. Report for government (P2)	18,624	Summer 2021	Co-authored; third author
3. Dissertation (P3)	18,824	Spring 2022	Single-authored

3.3.1.3 Adam

Three papers were collected from Adam, all assignments from two courses in his first semester of PhD studies. Both courses were taught by the same instructor. One of the three papers was a book review (P1), the other two the final projects from the two courses, where he submitted two complete papers with data analysis and results: one about university websites (P2), the other about a social media platform (P3). P1 and P2, specifically, were assignments of one course, and P3 was the final project of another course. The book review was about a dissertation which focused on a Chinese website. Before writing the book review, one of the course assignments was for the students to make a comparison among three book reviews (This piece of assignment was also collected for this dissertation study as part of the supplementary materials). Therefore, from the comparison, Adam could identify the features of book reviews. Based on that assignment, Adam then worked on his own book review. So, his own version was influenced by and constructed on the relevant book reviews he read in that course. Data analyzed in P2 and P3 were also related to Chinese context: P2 examined university websites that were located in China, whereas data for P3 were collected from a Chinese social media platform. Additional data included the course syllabi, and the instructor's comments on P3. Information about Adam's papers is presented in Table 6.

Table 6. Description of Adam's papers

Types of papers	Word count	Dates completed	Descriptions
1. Course assignment - Book review (P1)	1,681	Spring 2022	Single-authored
2. Course paper - Paper about university websites (P2)	7,255	Spring 2022	Single-authored
3. Course paper - Paper about a social media platform (P3)	6,661	Spring 2022	Single-authored

Note that the word count in the dataset was reached after excluding abstracts, references, appendices, and table of content – It only covered the main text.

3.3.2 Interviews

Interviews are the most common type of technique in qualitative studies (Dörnyei, 2007), as they can produce qualitative data that cannot be attained from other research methods (Bernard & Ryan, 2010). Interviews are also found to be a research tool that can open up space for students' academic progress (e.g., Morton et al., 2015). In this study, interviews were a primary way of collecting participants' background information, insights, and perspectives concerning identity constructed from their papers. It concerns in this specific research context, how discoursal identity (*discoursal self* and *self as author)* is shaped and influenced by context factors (*autobiographical self* and *possibilities for selfhood*). In other words, how writer identity construction is mediated by contextual factors (i.e., discourse, and genre) and individual factors (i.e., agency) were explored. I adopted semi-structured interviews because it allowed me to ask follow up questions when participants offered some answers that were worth probing. This in turn made the "purposeful conversation" (Richards, 2009) less rigid and rule-based and added a conversational tone to the interviews, which potentially made the participants more comfortable and less stressed.

More specifically, the interviews conducted included: 1) background interviews, which aimed to elicit students' prior educational background and their writing experience (their *autobiographical self*); 2) interviews about the students' papers, which were conducted based on students' papers after they were analyzed. Text features and students' reasons, beliefs, and perceptions of the discoursal identity projected in the papers were explored. Further, Chinese was used when I did interviews with the three Chinese doctoral students. Before meeting, I would contact them and select a time slot that worked for both of us. Meeting face to face was preferred, but when meeting in person was not possible, the interviews were conducted virtually. Interviews were audio-recorded. The information about the interviews with the three participants is displayed in Table 7.

Table 7. Information of the interviews

Participants	Semi-structured interviews	Dates conducted	Modes	Durations
Momo	Background interview (BI)	11/27/2021	In person	72 Minutes
	Interview on transportation roundabout paper - P1 (I1)	2/20/2022	In person	69 Minutes
	Interview on survey paper – P2 & P3 (I2)	11/5/2022	Virtually	64 Minutes
Nancy	Background interview (BI)	11/30/2021	In person	98 Minutes
	Interview on journal manuscript – P1 (I1)	2/10/2022	Virtually	152 Minutes
	Interview on government report – P2 (I2)	4/16/2022	Virtually	73 Minutes
	Interview on dissertation – P3 (I3)	10/29/2022	Virtually	58 Minutes
Adam	Background interview (BI)	3/16/2022	In person	57 Minutes
	Interview on book review – P1 (I1)	4/19/2022	In person	45 Minutes
	Interview on university websites paper – P2 (I2)	6/16/2022	In person	53 Minutes
	Interview on social media platform paper – P3 (I3)	9/24/2022	Virtually	34 Minutes

3.3.2.1 Background interviews

The background interviews concern student writers' *autobiographical self*. The interviews were conducted prior to textual data collection for the purpose of eliciting participants' basic background information, including their learning experience, English proficiency level, academic writing experience, writing instructions they received, research foci, academic goals, perceptions of writing and English language writing, and difficulties with academic writing. The information collected from the preliminary interviews were used to create a profile of each participant, as presented in the previous section. Also, participants' basic information scaffolded the subsequent paper interviews, which made it possible for me to solicit their ideas concerning academic writing and identity construction in a more contextualized manner. For example, Nancy used to work part time as a TOEFL teacher in China, and in the background interview, she mentioned that her colleague, a GRE writing teacher, provided her with valuable writing tips, which helped her tremendously in her writing papers in the PhD studies. In the subsequent paper interviews, I asked her more specifically how, if any, the writing tips informed her writing the papers she sent me; This shows the intrinsic relationship between *autobiographical self* and discoursal identity. The background interview guidelines can be found in Appendix A.

3.3.2.2 Interviews about papers

The second set of the interviews were conducted after I did the analysis on the participants' textual materials. It aims to investigate 1) what contextual factors (*possibilities for selfhood*) including context, and individual factors (i.e., agency) shape the construction of identity (*discoursal self* and *self as author*) in the participants' writing. Questions were asked about the papers which aimed to prompt the participants to reflect on and talk about what the reasons behind the choice of certain metadiscoursal markers that contributed to their identity construction.

Since interviews on papers revolved around identity in their writing, and contextual and individual factors influencing identity construction were negotiated during their writing process, two sets of questions were asked: The first one concerned more generally the requirements for their writing assignments, the characteristics of successful writings, the goal of writing, the main focus in writing, their attitude towards teacher feedback, the impressions their writing left on other people, their perceptions of their own writing, and the aspects of their writing skills needed improvement. The second part of the interviews was mainly based on metadiscourse use in their writing. The participants were interviewed with questions such as the purpose of using certain metadiscourse markers, the role the target audience played in impacting the final writing product, the kinds of effect they wished to achieve through writing, and their level of awareness of the impact of the metadiscourse markers on readers.

More specifically, for the metadiscourse markers, before the interviews, I first analyzed my participants' papers, and paid attention to the metadiscourse markers they tended to use. For instance, in Nancy's dissertation (P3), I noticed she provided explanations for a statement in parentheses extensively throughout the dissertation. Then in the interview, I consulted her about the reasons for doing so. I also asked questions based on what I learned of my participants in the background interviews. For instance, when I interviewed Nancy about the *attitude markers* used in her dissertation, I brought up her IELTS learning experience and further asked her if that was related to her use of those words expressing her affective stance. Meanwhile, I paid attention to "anomalies" that appeared in my participants' papers in relation to metadiscourse use, and in the interviews, I consulted their reasons for using them. For instance, in Momo's P2, the paper submitted for a conference, she used a colloquial sentence "The jury is still out" to express her uncertainty. Considering this was an interesting linguistic phenomenon, I asked Momo for the reasons for selecting this expression in the interview (I2). The analysis will further be reported in the results chapter. The interview questions about the papers can be found in Appendix B.

Note that a follow-up interview (FI) with Momo was conducted to seek for more information not very clear from the previous interviews.

3.4 Data analysis

In order to investigate each research question, different methods of data analysis were used. For the first research question examining what kinds of identities are constructed through metadiscourse, and what metadiscoursal patterns can be found, textual analysis was conducted by applying metadiscourse framework (Hyland, 2018) and the interactional model (Hyland, 2005) to the analysis. For the second research question, thematic analysis on the interview data was used to explore how and why the writers' identities were shaped the way they were, and the reasons for choosing certain metadiscourse markers to construct their identities influenced by context and personal agency.

3.4.1 Analysis of identity markers: metadiscourse

The analysis of identity markers through metadiscourse is to address the first research question. Due to the contextualized nature of metadiscourse, the data were analyzed manually in order to more precisely locate the linguistic chunks that attain metadiscoursal functions (Lee & Casal, 2014). The textual data were analyzed using MAXQDA 2022, a qualitative and mixed methods data analysis software (MAXQDA Product). Since it is noted that metadiscourse is a fuzzy category and its analysis is hard to apply in practice (Hyland, 2017), there were a few considerations when analyzing the data. First, some linguistic items may fulfill propositional functions instead of metadiscoursal functions (e.g., Hyland & Tse, 2004), as they refer to or connect propositional content outside the text instead of working internally within the text to fulfill the metadiscoursal functions. That is to say, they signal or reflect externally the relationships in the real world rather than function to serve the textual purposes internally. Therefore, those linguistic markers achieving propositional functions were not included in the analysis. For example, in Adam's social media platform paper (P3) in the following example, the "aimed at" actually refers to something existing outside the text in the real world – the goals of a social media platform. It is not internally associated with announcing the goals his paper aimed to address. And "Firstly" does not frame the text structurally but refers to the initial goals when the founder established the website. Therefore, even though the two linguistic segments can be found in previous lists of metadiscourse markers (e.g., Hyland, 2018, Lee & Casal, 2014), they were not counted as metadiscourse in this dissertation.

> Xionghongshu (lit. Little Red Book) was established by Wenchao Mao in 2013. <u>Firstly</u>, it <u>aimed at</u> constructing a cross-border e-commerce platform for individuals who were interested in overseas shopping. (Adam - P3).

Second, various metadiscoursal functions can lump together in a string of text, with larger units encompassing smaller units achieving multiple metadiscoursal functions. However, there are different ways to code the linguistic units – either coding it altogether as a whole unit or coding individual units one by one into its corresponding metadiscourse category (Hyland, 2017). For instance, in Momo's survey paper (P3), she indicated her attitude by using "important", and "particularly" was used before "important". Instead of treating the phrase performing one metadiscoursal function, this dissertation categorized them into two functions: *booster* and *attitude marker*.

> … during the test ride phase is <u>particularly</u> <u>important</u> for the attitudes of high-income people. (Momo – P3)

Procedures of data analysis using MAXQDA are as follows: First, the metadiscourse categories were the codes assigned to the linguistic segments in the papers realizing the metadiscoursal functions. The codes were created in the software, then the data were imported to MAXQDA 2022. The linguistic units that realized metadiscoursal functions were labeled with the corresponding metadiscourse categories in the software. The analysis only included the main text as the data; abstract, footnotes and appendices were excluded from analysis. Direct quotations, translations, interview quotes in the main texts were not included in the analysis either. Given the different lengths of the participants' papers, results are presented after being normalized per 1,000 words to ensure comparison within and across cases.

Intercoder reliability was established with a second coder who has research experience with metadiscourse coding three papers written by the three participants, approximately 25% of the data. The second coder was a PhD student from applied linguistics whose area of focus was related to second language writing, and metadiscourse was one of the foci among the research she did. Further, she and I have coded metadiscourse markers together for my previous project about two years ago. During that time of collaboration, we achieved our common understanding of metadiscourse. The previous experience working as raters had greatly facilitated our coding for this dissertation project. The intercoder agreement reached 99% after discussion. Given the fact that the second coder and I only had less than ten cases in total in disagreement, and metadiscourse is a fuzzy category (Crompton, 1997; Hyland, 2017), it can be considered acceptable to not reach 100% full agreement. After ensuring intercoder reliability, I coded the rest of the data on my own. During the coding, I took notes on those strings of words that I could not determine exactly which metadiscourse category they belonged to. After completing analyzing all the textual data, a few weeks later, I doublechecked my initial coding to ensure its accuracy. I paid special attention to the notes I made, and re-coded those words that was initially ambiguous to me.

3.4.2 Analysis of interview data

The interviews on the papers elicited the doctoral students' perceptions of writer identity and why they projected the identity the way they did in their papers. The analysis of interview data aimed to address the second research question. Thematic analysis were employed to analyze interview data. Thematic analysis refers to the process of encoding qualitative information (Boyatzis, 1998), i.e., interviews. It is a process of qualitatively analyzing interviews and identifying themes that emerge from interviewees' answers.

More specifically, first, the interviews were transcribed using an online tool (https://app.xunjiepdf.com/voice2text/). The rough transcriptions were corrected by the author and sent to participants for review. Each participant's writing samples and interview data were saved in separate files. In order to link data in dense, well-developed, and related categories, the coding consists of three phases: open coding, axial coding, and selective coding (Strauss & Corbin, 1998). Open coding is the initial and fundamental way of analyzing the data, which primarily aims to identify concepts, their dimensions, and properties. After finishing transcribing the recordings, I read through students' interview transcripts a few times. According to Merriam (1988), "[a]t this stage, the researcher is virtually holding a conversation with the data, asking questions of it, making comments, and so on" (p. 131). During the process, I took down notes, comments, and observations. Researcher's initial meaning-making of the interview data will then be "developed into a primitive outline or system of classifications into which data are sorted" (Goetz & LeCompte, 1984, p. 191). During this stage, I generated descriptive codes based on the ideas shared by my participants, primarily to capture the main topics of the interviews. For example, Momo talked about her being a writer who received help from her advisor in China regarding various aspects of her writing. I labeled codes such as "advisor's influence on self-mentions" and "advisor's focus on feedback - logic" to the corresponding statements.

After that, axial coding was started to "begin the process of reassembling data that were fractured during open coding". It is an act of "relating categories to subcategories along the lines of their properties and dimensions" (Strauss & Corbin, 1998, p. 124). This process encompasses looking for associations and recurring regularities among the codes in order to derive more inferential and explanatory pattern codes (Duff, 2008). At this stage, I was looking for how codes link, paying attention to salient or rival patterns, and the presence of anomalous or unforeseen things in the data, and examining the potential relationship between the data and theory as well as any inconsistencies and contradictions in the data (Hammersley & Atkinson, 1983; Yin, 2018). In this stage, the codes generated from open coding were grouped together to their corresponding categories. For instance, the descriptive codes mentioned in the last paragraph were categorized further to "advisor".

The next step is selective coding, which aims to integrate the categories with the sub-categories at a conceptual level, and refine and develop them into theory that has explanatory power within the context of the research (Moghaddam, 2006). Merriam (1988) notes that "[d]evising categories is largely an intuitive process, but it is also systematic and informed by the study's purpose, the investigators' orientation and knowledge, and 'the constructs made explicit by the participants of the study'" (p. 133). Bearing this in mind, when creating the conceptual categories, I was guided by the conceptualizations of writer identity from the social constructivist perspective, and highlighted the themes related to the constructs that influenced writers' identity construction (i.e., context and agency). For instance, mapping onto the conceptual framework adopted in this dissertation, the code "advisor" generated from axial coding was mapped onto *possibilities for selfhood,* since it concerned the external force that contributed to Momo's identity construction. The codes and categories generated could be exemplified in Figure 5.

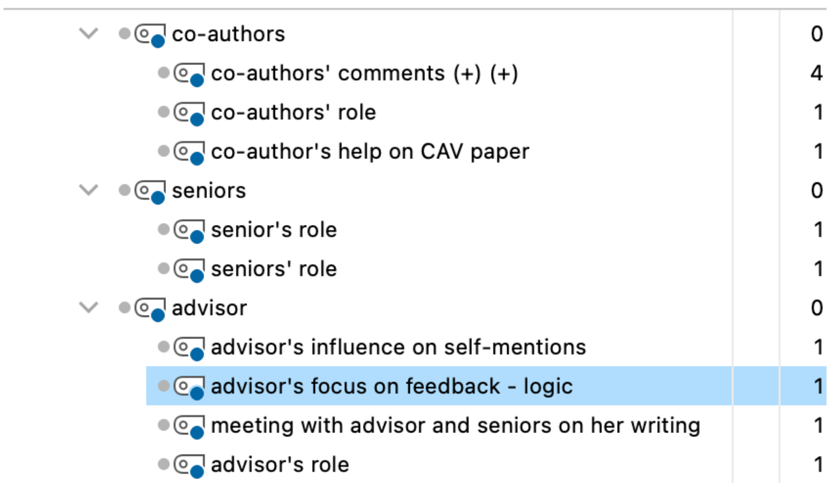

Figure 5. An example of interview data analysis using MAXQDA

Data analysis in qualitative case study is *iterative*, *cyclical*, and *inductive* in nature (Duff, 2008). Therefore, the above process was not linear but went through multiple rounds of revising, synthesizing, organizing, and combining. Also, the data analysis process was guided by the theoretical conceptualizations of writer identity; and in turn, the insights gained from the data can scaffold the conceptual interpretations of the constructs concerned in the study.

After the analysis on interviews was done, I conducted member check to ensure the accuracy and completeness (Gall et al., 2005) of data analysis on the interviews to enhance the trustworthiness of the study. I first completed writing the results corresponding to each participant, then the part of results that were generated from analysis of interviews (results to RQ 2) were sent back to my participants to read. I kindly asked them to check if any of my understandings were incorrect and if they had anything they wanted to add. All the three participants agreed to my understanding and interpretations. Nancy specifically remarked that "You didn't misunderstand me. This is a true record of fact". Momo, after reading my interpretations, preferred a pseudonym than the one I initially chose. So "Momo" was a name she chose for herself after she read my analysis on her data.

3.4.3 Analysis of supplementary data

The supplementary data, as they can also provide additional information, were coded in order to scaffold the analysis of the textual data and interviews, and provide additional insights. The supplementary materials were coded based on the information they revealed related to metadiscourse use and other constructs in this study, including *autobiographical self, possibilities for selfhood*, and agency. For instance, in Momo's research log, a huge amount of Chinese was used, which could provide support in regard to how native language (aka. her being a Chinese person) influenced her identity construction in papers. This was reported in the results chapter as Momo's *autobiographical self* influencing her writer identity in her papers. In the feedback Adam received for his P3, I focused on the comments provided by the instructor that were related to metadiscourse use. For example, a few comments specifically mentioned the citation use, which showed how instructor's feedback shaped Adam's identity construction.

Instructor's feedback, as an external source of influence, was categorized into *possibilities for selfhood* as a factor contributing to Adam's identity construction.

3.5 Trustworthiness of the study

It is hard, if not impossible, to separate the assessment of trustworthiness from examining a study's components (Merriam, 1988). This study has employed several methods to ensure the trustworthiness. First, triangulation is seen as a reliable way to ensure the trustworthiness of a qualitative study, which can employ "multiple methods, data sources, and researchers to enhance the validity of research findings" (Mathison, 1988, p. 13). Looking to provide an analysis on L2 doctoral student writers' identity construction in papers of different kinds, this study employed data and method triangulation to enhance the validity and trustworthiness of the study (Flick, 2018). As to data triangulation, it included writers' writing assignments from different periods, interview data with student writers, and various materials such as syllabi, and journal reviewers' comments to provide additional insights to the issue under investigation. Methods triangulation included interviews and textual data analysis together with the analysis of other documentations. Further, intercoder reliability regarding the analysis of textual data was obtained, and member check regarding the interview data were conducted to make sure my accurate understanding of the participants' ideas (Yin, 2018).

3.6 Summary of the chapter

In this chapter, I presented the methodology employed in this dissertation research to answer the research questions. I first discussed the rationales for adopting a qualitative multi-case study design, then moved on to the descriptions of the context and the selection of my participants. Next, I elaborated on the data used in this study, including textual data and interview data, and other supplementary materials. Textual data included three participants' three writing samples of different kinds, and interview data encompassed the background interviews I did with them prior to data collection, and interviews on their papers. Following this, I presented the analysis of each type of data.

Table 8 summarizes data collection and data analysis for investigating each research question. The first research question "How do L2 doctoral students construct writer identity in their papers of different kinds?" entails two sub-questions. The first one "What metadiscoursal features does their writing have that contribute to the construction of writer identity?" aims to look into the metadiscoursal features in the L2 doctoral students' writing and addresses the *discoursal self* and *self as author* formulated from the texts. Textual analysis was used to answer this research question. The second part of the first research question asks "What are the patterns of metadiscourse use in identity construction in the doctoral students' papers across different writing tasks?" aims to unravel the similarities and differences of identities constructed across the doctoral students' three papers exhibited in the use of metadiscoursal markers. Textual analysis on metadiscourse resources was adopted to answer this research question. The second question "How do contextual factors and agency shape the doctoral students' construction of identity in their papers?" investigates the role that *autobiographical self* and *possibilities for selfhood* (i.e., discourse, context) and student writers' agency play in shaping their discoursal identity. The data for analysis for this question are interviews with the students, their papers as well as supplementary materials including paper reviewers' comments and revisions.

Table 8. Data collection and data analysis by research question

<table>
<tr><th colspan="2">Research questions</th><th>Aspect/s of writer identity to be examined</th><th>Related construct/s</th><th>Data collection</th><th>Data analysis</th></tr>
<tr><td rowspan="2">1. How do L2 doctoral students construct writer identity in their papers of different genres?</td><td>1a: What metadiscoursal features does their writing have that contribute to the construction of writer identity?</td><td>Discoursal self and self as author</td><td>Metadiscoursal resources</td><td>L2 doctoral students' papers</td><td>Textual analysis</td></tr>
<tr><td>1b: What are the patterns of metadiscourse use in identity construction in the doctoral students' papers across different written genres?</td><td>Discoursal self and self as author</td><td>Metadiscoursal resources</td><td>L2 doctoral students' papers</td><td>Textual analysis</td></tr>
<tr><td colspan="2">2. How do contextual factors and agency shape the L2 doctoral students' construction of identity in their papers?</td><td>Autobiographical self and possibilities for selfhood; Discoursal self and self as author</td><td>Context; agency</td><td>Background interviews and paper interviews with doctoral students

Supplementary documents</td><td>Textual analysis; Thematic analysis</td></tr>
</table>

CHAPTER 4: FINDINGS

In this chapter, I report the findings of this dissertation study. I present the cases starting from Momo, followed by Nancy and Adam. Within each case, I start with the linguistic distribution of metadiscoursal markers constructing the doctoral student's identity to answer research question 1. To answer research question 2, I make a comparison across the different types of papers and examine the patterns through metadiscourse usage in regard to identities constructed in the doctoral writer's papers. To answer research question 3, I address reasons, including contextual (*autobiographical self* and *possibilities for selfhood*) and individual agency, which influenced the doctoral writers' identities (*discoursal self* and *self as author*) constructed in their papers. Last but not least, after presenting the results of each case, I make a comparison across the three doctoral students and highlight the similarities and differences in regard to their identity construction in their academic papers.

4.1 Momo

4.1.1 Metadiscourse usage in identity construction in Momo's papers

In this section, I report how Momo's identities were constructed through the use of metadiscourse, starting with the overall distribution of metadiscourse in Momo's papers. Then I report the distribution of each category of metadiscourse in the sequence of interactive metadiscourse, and stance and engagement markers under interactional metadiscourse.

4.1.1.1 Overall distribution of metadiscourse markers in Momo's papers

As displayed in Table 9, in total there were 761 cases of metadiscourse in Momo's three papers. 514 cases of interactive metadiscourse were found, which occupied 67.5% of the metadiscourse markers, 22.5 cases per 1,000 words. 229 times of stance markers were used in the three papers, which covered 30.1% of the total markers, and 10 cases per 1,000 words. Engagement markers were used the least frequently, appearing 18 times in total, covering 2.4% of the total metadiscourse occurrences.

In Momo's first manuscript for publication (P1), 222 cases of metadiscourse markers were found, which was 31.7 times per 1,000 words after being normalized. More specifically, 159 cases of interactive metadiscourse items were found, covering 71.6% of the total metadiscourse. Stance markers occurred 59 times in this paper, constituting 26.6% of the total metadiscourse resources. Engagement markers, however, were used minimally in Momo's P1, appearing only 4 times, covering 1.8% of the metadiscourse use in this paper.

In Momo's paper submitted for conference presentation (P2), 204 cases of metadiscourse were found, which equaled to 26.8 times per 1,000 words after normalization. Among the metadiscourse items, interactive metadiscourse was used the most frequently, which appeared 140 times, covering 68.6% of the total metadiscourse markers. Stance markers, on the other hand, were used 61 times, covering 29.9% of the total metadiscourse markers in P2. Engagement markers were used to a considerably lesser extent: only 3 cases were found, 0.4 times after being normalized per 1,000 words.

Another manuscript for publication (P3) of Momo's showed a use of 335 instances of metadiscourse markers, which amounted to 40.9 times per 1,000 words. In this paper, 215 instances of interactive metadiscourse were used, covering 64.2% of the metadiscourse markers.

109 stance markers were found, which constituted 32.5% of metadiscourse markers. The least frequently used metadiscourse category is engagement: only 11 cases were found, covering 3.3% of the metadiscourse markers in P3.

Table 9. Overall distribution of metadiscourse markers in Momo's papers

	P1		P2		P3		Total	
Category	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words
Interactive metadiscourse	159 (71.6%)	22.7	140 (68.6%)	18.4	215 (64.2%)	26.2	514 (67.5%)	22.5
Interactional metadiscourse (stance)	59 (26.6%)	8.4	61 (29.9%)	8	109 (32.5%)	13.3	229 (30.1%)	10
Interactional metadiscourse (engagement)	4 (1.8%)	0.6	3 (1.5%)	0.4	11 (3.3%)	1.4	18 (2.4%)	0.8
Overall total	222 (100%)	31.7	204 (100%)	26.8	335 (100%)	40.9	761 (100%)	33.3

4.1.1.2 Interactive metadiscourse and identity construction in Momo's papers

Table 10 displays the interactive metadiscourse used in Momo's three papers. Interactive metadiscourse functions to link propositions, offer elaborations and present supporting materials. It helps organize a text for readers to "recover the writer's intentions, creat[e] surface cohesion, and influenc[e] understandings of propositional material" (Hyland & Jiang, 2022, p. 8). That it was dominantly used in Momo's three papers suggested Momo's efforts to glue texts together and to anticipate readers' needs and to guide readers through the reasoning process, so that readers could keep informed about what has been specified and what yet to be offered.

Transitions were used the most frequently among the five sub-categories of interactive metadiscourse across Momo's three papers. As an important academic writing feature (Hyland, 2018), *transitions* could specify the logical relationship between ideas, making the connections of information comprehensible to readers. For example,

> … given the fact that such a control strategy shall be updated at the frequency of seconds, while optimization-based strategies typically require minutes to complete their computation (Wang et al., 2019). In addition, their computation time increases… (P1)
>
> Previous research has only considered the population group that is prone to change their attitudes (*9*, *12*) but has not examined what causes these changes. Moreover, the results of the test ride are no guidance on how to improve the test ride. (P2)
>
> Seven variables are found to affect participants' opinion change on AV safety significantly. Therefore, the estimated utility functions are:… (P3)

The employment of *transitions* in the instances above showcased Momo's understandings of the logical relations in the texts, which also helped readers to comprehend and interpret the texts. The use of the *transitions*, the explicit marker of signaling textual connection, could demonstrate her identity as a writer who was conscious of the academic writing conventions and who organized her texts using the transitioning words indicating conspicuous logic.

Table 10. Interactive metadiscourse in Momo's papers

	P1		P2		P3		Total	
Category	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words
Transitions	49 (22%)	7	49 (24%)	6.4	71 (21.2%)	8.7	169 (22.2%)	7.4
Frame markers	15 (6.8%)	2.1	11 (5.4%)	1.5	18 (5.3%)	2.2	44 (5.8%)	1.9
Endophoric markers	34 (15.3%)	4.9	19 (9.3%)	2.5	30 (9%)	3.6	83 (10.9%)	3.6
Evidentials	38 (17.1%)	5.4	36 (17.6%)	4.7	60 (18%)	7.3	134 (17.6%)	5.9
Code gloss	23 (10.4%)	3.3	25 (12.3%)	3.3	36 (10.7%)	4.4	84 (11%)	3.7
Interactive (Total)	159 (71.6%)	22.7	140 (68.6%)	18.4	215 (64.2%)	26.2	514 (67.5%)	22.5

The second most frequently used interactive metadiscourse was *evidentials*. Including citational support in papers is considered a well-acknowledged writing convention in academic written discourse, which could situate one's research in the web of knowledge constructed by the broader academia. For instance,

> … their computation time increases exponentially with the increase in the number of vehicles (Chen and Englund, 2016; Meng et al., 2017; Müller et al., 2016). (P1)
>
> These factors include demographic characteristics (*5, 7–10*), driving habits (*7*), psychological characteristics, and AV-related experience (*9, 11–14*). (P2)
>
> Current research also found that public perception is also influenced by demographic features (Deb et al., 2017; Hulse et al., 2018; Pyrialakou et al., 2020) and favorable interpretations of the brand of AV (Reig et al., 2018). (P3)

Mainly used in the introduction and literature review sections, *evidentials* appeared rather frequently in Momo's papers. By drawing on previous knowledge and integrating the known in her current papers, Momo not only strengthened the persuasive power and the credibility of the content in the papers, but also demonstrated herself as a conscientious researcher who had read extensively relevant literature and synthesized previous research in her paper. *Evidentials* chosen by Momo also demonstrated the discourse community that she identified herself with (Shi, 2011).

Code gloss was another strand of interactive metadiscourse that was frequently used across Momo's three papers. *Code gloss* represents a range of linguistic devices that provide further explanations and reformulations of materials. In Momo's papers, *code gloss* was used extensively to fulfill different kinds of rhetorical functions. Momo introduced plenty of formulas in the papers, and how the technical knowledge worked and how it was applied to her study needed detailed explanations (such as the example in P1 below), which may require the deployment of *code gloss*. Further, she also offered more specific and elaborate information about certain phenomena (examples in P2 and P3) to help readers process the information in an unambiguous way.

> $T_{0-1}^{w,k}$ <u>represents</u> the first acceleration part. $T_{1-2}^{w,k}$ <u>represent</u> the part vehicle driving at the maximum speed. $T_{2-3}^{w,k}$ <u>represents</u> the last part vehicle decelerating to v_r. (P1).
>
> Most of these studies are based on simulation experiments (*19*), <u>i.e., simulating various HV-AV interaction scenarios…</u> (P2)
>
> The carrying capacity of each AS was ten passengers (<u>8 seats and two extended seats</u>). (P3).

Using *code gloss* in the three papers constructed Momo's identity as an academic writer who was equipped with knowledge to offer explicit explanations to readers. It also showed that Momo as an academic writer, could anticipate readers' reactions and thus could address those concerns that readers may have.

Endophoric markers were found to be used rather frequently in Momo's three papers. They were used more in P1 than in P2 and P3. Specifying the structure of an article helped connect different parts of the text together, and it could prepare readers for what Momo was going to present, which would help readers process the information more effectively (such as the example extracted from P1 below).

> The remainder of the paper is organized as follows: <u>Section 2 describes the problem considered in this study; Section 3 describes the proposed controller and its solving process; Section 4 conducts simulation evaluation and sensitivity analysis. Finally, conclusions and recommendations are delivered in Section 5.</u> (P1)
>
> <u>Figure 1</u> shows the AS operated by Pinellas Suncoast Transit Authority (PSTA). <u>Figure 2</u> shows the route of the AS. (P2)
>
> The detailed questionnaire form used for collecting the data is provided <u>in the Appendix</u>. (P3)

It is a common practice that Tables, Figures, and graphs are used extensively to contribute to the multimodal presentation of information in the hard science field. In Momo's papers, they were used together with linguistic resources to depict a full-fledged description of the information presented in the Tables, Figures, and graphs. For example, in the P2 and P3 extracts above, Momo referred to other parts of the papers so that readers could locate the information more directly and precisely and connect the linguistic descriptions with visual presentation of the information.

The use of *endophoric markers* constructed Momo as a capable writer who could organize the texts together in an internally logical way, and a considerate writer who showed her awareness of writing conventions in her discipline, and could take into readers' processing efforts into account and specify the connections between materials directly.

Frame markers, on the other hand, were the least frequently used interactive metadiscourse in Momo's papers. They mainly functioned to explicitly structure a paper. As Momo's papers were full-length research articles, it was considered reasonable to use these signposts to guide readers through the reading journey, and to potentially facilitate reading process. *Frame markers* can be divided into four categories: announcing goals, sequencing, labelling stages, and shifting topics (Hyland & Jiang, 2022). The first three were found in Momo's papers. Announcing goals, by its name, refers to writers making clear the objectives of their study, as shown in the following example:

> <u>The primary goal of this study</u> is to investigate the factors influencing the public attitude toward AV shuttles and attitude change… (P3)

Pointing out the objectives of a study was crucial in academic papers, which would convey to readers the central issue Momo's papers intended to investigate.

Second, *frame markers* that were used to sequence a complicated entity were also found in Momo's papers. This would be highly useful when more than one point about a phenomenon or an argument need to be highlighted. In the following instance, Momo broke down what "the improvement" entailed in an approachable way by enumerating the two points it entailed. Explaining a concept this way helped disentangle a potentially complicated issue, contributing to readers' understandings:

The improvement can be explained in two folds: Firstly, it is a free-flow state when the conflicting flow rate equals 0… Secondly, it is an impeded-flow state when exceeds zero. (P1)

Third, Momo's papers contained *frame markers* for the purpose of labeling stages. The example below from P2 used "to summarize", a phrase indicating dialogical closure (Hyland & Jiang, 2018). This sentence appeared at the end of literature review when Momo had finished reviewing relevant literature and needed to offer the insights she generated that could be used to inform her own study.

To summarize, earlier studies ignored considering the influence of both demographic and interaction experience on attitudes… (P2)

Frame markers for different purposes were found in Momo's three papers. They served as important signposts gluing texts together, yet they simultaneously constructed Momo as a writer who could guide readers through the discussion using the explicit linguistic markers and assist readers to comprehend materials of her own explanation.

4.1.1.3 Stance markers and identity construction in Momo's papers

Table 11 demonstrates the distribution of stance markers in Momo's three papers. The four sub-categories of stance markers were found in all three papers except in P1, where *self-mentions* were absent. Another pattern is noticeable across the three papers: *hedges* and *boosters* were used significantly more than *self-mentions* and *attitude markers*.

Hedges were the most frequently used stance marker in Momo's P2 and P3, while in P1, they were used less than *boosters. Hedges* encompass linguistic devices that can alleviate writer's tone of assertiveness and certainty, which constructs propositions as personal opinions that can be open to readers' interpretations instead of indisputable facts (Hyland, 2018). Momo tended to use *hedges* frequently when she tried to offer explanations toward a research result. For instance:

> This value suggests that it is not significant that female participants have different parameter values… (P2).
>
> There is a chance that an earlier special temporal resources is not significant enough… (P1).
>
> This is probably because when pedestrians are involved in a crash with AS… (P2).
>
> … people with high incomes and long commute times are more seemingly to change their minds positively… (P3).

In the examples, Momo used words and phrases to express her uncertainty in making these propositions, which were mainly her personal opinions. Making a claim is risky as it may be in conflict with extant literature or challenge readers' perspectives (Hyland, 2018); hence to avoid oppositions, anticipating readers' reactions and downplaying the personal stamp on claims would be a useful strategy. By leaving space for alternative views or interpretations, Momo constructed herself as a prudent and careful writer who conveyed deference to readers to achieve appropriate academic persuasion.

Compared to *hedges*, *boosters*, on the other hand, indicate otherwise. *Boosters* are linguistic expressions writers draw on to strengthen their claims and emphasize the truth of a statement (Hyland, 2018). *Boosters* were also a rather frequently used stance marker in Momo's papers, and was used the most frequently in P1, but ranked the second in P2 and P3. *Boosters* signal writers' certainty and can strengthen an argument. Momo used *boosters* primarily to indicate that her statement was based on precise calculation; therefore, there was objective evidence to support Momo to make claims in such a confident and self-assured way. For example:

> The computational time for each optimization is always less than 0.02 sec across all traffic demands. (P1).
>
> Therefore, the throughput totally depends on saturated headway *h.* (P1).
>
> This result demonstrates that autonomous technology's performance in the early stages… (P2).
>
> This also shows that as AV shuttle technology involves more people trying its service… (P3)

In the examples, based on mathematical calculation, Momo was able to generate a fixed number, based on which she could draw conclusions self-assuredly. By boosting the propositions with emphatic verbs and/or amplifying adverbs (Lee & Casal, 2014), Momo constructed an identity of a self-confident writer whose assurances were constructed through data-supported reasoning, which was not prone to alter if the procedures were carried out by a different researcher.

Table 11. Interactional metadiscourse in Momo's papers – stance

	P1		P2		P3		Total	
Category	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words
Hedges	18 (8.1%)	2.6	26 (12.8%)	3.4	48 (14.3%)	5.9	92 (12.1%)	4
Boosters	34 (15.3%)	4.8	23 (11.3%)	3	42 (12.5%)	5.1	99 (13%)	4.3
Self-mentions	0 (0%)	0	6 (2.9%)	0.8	10 (3%)	1.2	16 (2.1%)	0.7
Attitude markers	7 (3.2%)	1	6 (2.9%)	0.8	9 (2.7%)	1.1	22 (2.9%)	1
Interactional - stance (Total)	59 (26.6%)	8.4	61 (29.9%)	8	109 (32.5%)	13.3	229 (30.1%)	10

Self-mentions were used only in P2 and P3, with 6 cases used in the former and 10 cases in the latter. It is worth mentioning that only "we" and "us" were used in Momo's papers, since, as stated in the methodology chapter, all three papers were co-authored. Using "we" and "us" refers to all of the authors, as shown in the following examples, where Momo used self-referenced pronouns to bring the writers to the fore:

> In this section, <u>we</u> briefly review the related works on three topics. (P2)
>
> The likelihood ratio test can help <u>us</u> determine whether different models are needed for different groups. (P2)
>
> The result shows <u>we</u> are over 95% confident in rejecting the hypothesis that the three opinions are the same. (P3).

Self-mentions are considered the most conspicuous linguistic resource that could signal writer's presence and their engagement with readers in academic writing (Hyland, 2001). With intrusion into the text using *self-mentions*, Momo constructed herself (along with the co-authors) as a writer who took responsibility for the claims and actions made in the article and who took credit for the interpretations (Hyland & Jiang, 2018). Though it is common that in hard disciplines (e.g., transportation engineering) personal roles tend to be downplayed to ensure that research process and results could be accomplished regardless of who conducts the study to reflect the replicability nature of research (Hyland, 2018), referring to writers indeed could help them accentuate their authorial roles as academic writers whose collective efforts were acknowledged in the topic under investigation and help them gradually build up academic credentials in their respective field (Hyland & Jiang, 2018).

Attitude markers were used the least frequently in the stance marker category, which covered 3.2%, 2.9% and 2.7% in P1, P2 and P3 respectively. *Attitude markers* convey writer's affective and evaluative judgements toward arguments or propositions under discussion. This category is not used very commonly since in academic writing there is relatively less context in need of the expression of affect (Biber et al., 1999). In Momo's three papers, her (and the co-authors') attitudinal stance toward the subject matters was apparent through the use of words such as "fortunately", "important", and "noteworthy".

> Fortunately, the Connected and Automated Vehicle (CAV) technology brings the possibility to further… (P1).
>
> This indicates that it is important to eliminate pedestrians' concerns. (P2).
>
> Among the six variables in the table that were significant across all age groups, there are some noteworthy differences. (P3).

In the field of engineering, though writers are restricted in their linguistic repertoire to write in a subjective way (Hyland, 2005), positioning personal and subjective stance through *attitude markers* has witnessed a gradual increase in the past 30 years (Hyland & Jiang, 2018). As Momo's papers showed, the use of the adverbial and adjectival words expressed her affective stance toward propositional assertions. That Momo expressed their judgments explicitly indicated their identity as authors who were informed with field knowledge and who could tease out what was worth highlighting and then make evaluations on their study in relation to the knowledge from the broader discipline. Using this stance marker also contributes to Momo's identity as an individual writer whose subjectivity was stamped on the propositions, making her personal voice more prominent.

4.1.1.4 Engagement markers and identity construction in Momo's papers

Table 12 shows the use of engagement markers in Momo's three papers, which were significantly less frequently used compared to the other two types of metadiscourse. This relatively low frequency indicated that Momo did not interact with or engage her readers extensively in her writing. The commonality shared across the three papers was that *questions* and *personal asides* were not used, but *directives* were found to appear in all of the three papers. Different from P1 and P2, where 2 and 1 cases of *directives* were found, in P3, there were 8 instances of *directives*. Hyland (2002a) proposed three activities *directives* instruct readers to do: *textual acts, physical acts*, and *cognitive acts*. Though *directives* were not used at a high frequency in Momo's papers, all three types were found. *Textual acts* are used to direct readers' attention to other parts of a text. In Momo's papers, there are usually Tables, Figures and formulas to present knowledge; this way, readers' attention was directed to particular part/s of the article through *textual acts* (often in brackets). For example:

> Given the time step is 0.2 sec (<u>see Fig. 7</u>)… (P1).

In this instance, using "see Fig. 7" not only helps connect the text in a coherent way, but also shows Momo trying to lower readers' cognitive processing load by directing readers' attention directly to the visual representation of the information.

The second type of directives, *physical acts*, was also found in Momo's papers. They were used mainly through assertive modal verbs, such as *must* and *should*. They were used predominantly in the conclusion section where Momo needed to point out the practical significance of the study and how the study could inform future practical implementation outside academia.

AV operators must take pedestrians' and drivers' concerns into account when promoting AVs, try to reduce mistrust… (P3).

Transit agencies should focus on rigorous assessment and validation… (P3).

Through instructing readers what to do in real world, Momo’s use of *directives* constructed herself as a writer who was confident of the significance of the research and her research outcome, which was based on precise calculation and computation.

Lastly, Momo used *directives* to fulfill *cognitive acts* mainly to highlight some aspects of her data, so that the readers could capture the intended meaning. She tended to use imperative sentences to achieve such rhetorical effect. In the following two extracts, by instructing readers the prominence of her data, Momo anticipated readers’ reactions and thus tried to direct readers’ processing the way she preferred. This demonstrated that Momo, as a writer, could manipulate linguistic resources to instruct readers to perform cognitive actions and see things in the way determined by herself (Hyland, 2002a; Hyland & Jiang, 2016).

Notice that a vehicle with a higher service priority does not mean it enters the roundabout… (P1)

Note that only one participant chooses "very uncomfortable". (P2)

Table 12. Interactional metadiscourse in Momo's papers – engagement

	P1		P2		P3		Total	
Category	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words
Reader pronouns	0 (0%)	0	1 (0.5%)	0.13	3 (0.9%)	0.4	4 (0.5%)	0.2
Directives	2 (0.9%)	0.3	1 (0.5%)	0.13	8 (2.4%)	1	11 (1.5%)	0.5
Questions	0 (0%)	0	0 (0%)	0	0 (0%)	0	0 (0%)	0
Appeals to shared knowledge	2 (0.9%)	0.3	1 (0.5%)	0.14	0 (0%)	0	3 (0.4%)	0.1
Personal asides	0 (0%)	0	0 (0%)	0	0 (0%)	0	0 (0%)	0
Interactional - Engagement (Total)	4 (1.8%)	0.6	3 (1.5%)	0.4	11 (3.3%)	1.4	18 (2.4%)	0.8

As to *reader pronouns*, P1 did not use this engagement marker, whereas they were found in both P2 and P3. In the very infrequent occurrence of reader pronouns, Momo exclusively used "we" to refer to the writers and engage readers. For example:

> We can observe that participants with long commute times intend to be conservative on AS comfort… (P2).
>
> … whereby we can examine the gap between the public's acceptance of AV… (P3)

Using "we" to include readers acknowledged author's awareness of the presence of readers as well as Momo's solidarity with the target readers. Her use of "we" also showed that bearing reader awareness in mind, Momo constructed herself as the agent of her paper, who simultaneously guided the readers through the reasoning, argumentation, and interpretations preferred in her own ways.

Lastly, *appeals to shared knowledge* were not found in P3, with 2 cases found in P1 and 1 in P2. In the two examples below, the employment of the expression "common" positioned Momo's identity as a writer who aligned herself with readers sharing the same disciplinary understandings to signal her disciplinary membership:

> … according to Poisson distribution, which is a common practice for traffic control at road nodes… (P1).
>
> The inability to deliver passengers from origin to destination is a common problem in the current transit system. (P2).

4.1.2 Patterns of identity construction through metadiscourse use in Momo's three papers

In this section, based on the results of last section, I compare the ways identities were constructed through metadiscourse use in Momo's three papers. I start with the overall pattern of the metadiscourse markers, followed by interactive metadiscourse markers and interactional metadiscourse.

4.1.2.1 Overall patterns of metadiscourse usage in Momo's papers

As demonstrated in Table 9, several patterns regarding metadiscourse usage were emergent from Momo's three papers: First, interactive metadiscourse was overwhelmingly used in the three papers, ranking the most frequently used metadiscourse category, followed by stance markers and engagement markers. Engagement markers were used negligibly in the three papers. This showcased that Momo put more efforts in gluing her texts together to make her texts function as a meaningful unity and focused comparatively less on the interaction and engagement with readers.

Second, the tendency of overall use of metadiscourse features and the sub-categories displayed the same pattern. As has been stated in the methodology chapter, Momo wrote the three papers at different stages of her PhD studies: P1 in the first semester, P2 at the end of first year, and P3 at the beginning of the second year. P3 was a revised version of P2 for the purpose of publication. The chronological sequence of the papers also demonstrated the developmental trajectory of metadiscourse usage in Momo's papers. Overall, use of metadiscourse markers and the subcategories of metadiscourse increased overtime: compared to P1, in P3, each category was used more frequently. However, in P2, overall metadiscourse and use of each sub-category declined in comparison to P1. The detailed patterns of each metadiscoursal category are reported in the following sections.

4.1.2.2 Patterns of interactive metadiscourse in identity construction in Momo's papers

As displayed in Table 10, in regard to interactive metadiscourse, there was a general pattern across Momo's three papers: interactive metadiscourse was used the most frequently in P3 after normalization; the least frequently used, however, was in P2. The five sub-categories of interactive metadiscourse, in a similar vein, demonstrated the same pattern. *Transitions, frame markers, endophoric markers, evidentials,* and *code gloss* were all used the most frequently in P3, while the least frequently in P2.

That use of each type of the interactive metadiscourse demonstrated a consistent pattern in Momo's construction of identity seemed interesting. The metadiscourse usage could be compared along two lines. First of all, the increasing trend from P1 to P3, the two papers of the same genre - the research article aiming for publication, showed Momo's identity construction trajectory diachronically. As described in the methodology chapter, P1 was a paper written in the first semester of Momo's PhD studies, whereas she wrote P3 in the third semester of her doctoral studies. More interactive metadiscourse use in P3 demonstrated that after one year of socialization into the target discourse community, Momo was able to construct herself as a more competent writer who could use signposts to connect text together and thus could more skillfully guide readers through the text. Second, that interactive metadiscourse was used the least frequently in P2 whose purpose was a conference presentation showed that identity construction was sensitive to different situational context and genres. Momo regarded the paper for conference presentation and the one for journal publication differently. She was very clear about the purposes of P2, which was to present at an annual conference in her field and not to get published. It can be concluded that compared to reporting ideas, organizing the text in a more reader-friendly way like manuscripts for journal publication may be a less imposing rhetorical purpose to attain for P2. This can probably explain why all metadiscourse markers were used the least frequently in P2 among Momo's three papers.

4.1.2.3 Patterns of stance markers in identity construction in Momo's papers

As to stance markers exhibited in Table 11, a less homogenous pattern can be found. Use of *Boosters* and *attitude markers* followed the same pattern as interactive metadiscourse – in P2 they were used the least frequently, whereas in P3 they were used the most frequently. This probably showcased that in P2, Momo's primary goals were not to get the paper published but only to present the essential information to conference attendees, which reduced the needs to position her as a reflective writer who integrated her own attitude into the final writing product.

On the other hand, *hedges* and *self-mentions* showed an increasing trend diachronically in Momo's three papers: the least frequently used in P1, the most in P3. That *hedges* were used increasingly frequently demonstrated Momo trying to construct identities as a cautious academic writer who gradually learned the skills of maneuvering linguistic choices to strategically augment her credibility. However, though *self-mentions* also demonstrated an increasing pattern, they were not found in P1, where "this paper" was used to downplay the presentation of authors, such as:

> Motivated by these research gaps, <u>this paper</u> presents an advanced control approach for roundabouts with three features:... (P1)

In P2 and P3, however, "we" occurred more frequently, showing the tendency of Momo's efforts of bring herself (and other co-authors) to the fore, which contributed to the construction of her identity as an authoritative academic writer.

4.1.2.4 Patterns of engagement markers in identity construction in Momo's papers

As displayed in Table 12, though engagement markers were used rather infrequently, several patterns could be visible from the data. First, Momo did not use *questions* or *personal asides* in any of the three papers. This may be caused by the disciplinary writing conventions, as in the engineering field, academic writing is not a highly dialogically engaging discourse (Hyland, 2002). This would make initiating *questions* and inviting readers to answer those questions a somewhat less preferred move. The commonly found questions in academic discourse are research question/s that a study aim to investigate. However, in Momo's papers, the research questions were often specified using declarative sentences. For example,

> Since it has been proven that AV ride experience has a positive effect on public attitude toward AV (*9*), it is necessary to study whether AV ride experience also has a positive effect on people's willingness to share the road with AV. (P2)

Further, compared to more discursive fields, inserting personal opinions through *personal asides* in scientific writing generally seems a less common strategy, and readers do not have to be drawn in and involved as participants in the texts in the hard fields (Hyland & Jiang, 2016). The only category that had been used in all three papers were *directives*, which were common to use by Momo mainly to make suggestions for future research and point out the significance of the research for practical use.

4.1.3 Factors shaping Momo's identity construction in the three papers

From a social constructivist approach, identity is constructed under specific sociocultural context as well as is susceptible to individual writers' choice (i.e., agency). In this section, I report factors that had a sway on Momo's identity construction embodied through metadiscoursal resources in her three papers.

4.1.3.1 *Autobiographical self* and *discoursal self* and *self as author*

Momo's *autobiographical self* (i.e., "life history") (Ivanic, 1998) has informed her construction of identities in the three papers. More specifically, it encompassed her real-world identity as a Chinese person and her role as an anonymous reviewer.

4.1.3.1.1 Being a Chinese person

As discussed in Momo's profile, Momo completed her master's degree in China. When working on the master's degree, she had extensively written academic papers in Chinese. This writing experience affected her way of writing as an ESL PhD student in the US. As a Chinese international student, Momo said, "When writing papers in English, sometimes I will write in Chinese first, like keeping a record of the experiment. Then I translate it into English. Or sometimes if I really don't know how to write, I will write in Chinese then translate" (BI). This was substantiated in Momo's research log, where Chinese was used extensively to take notes. In the following figure clipped from her research log for P2 and P3 (the translations are provided on the right side), Momo wrote down the considerations that should be made in writing discussion and conclusion sections. More specifically, she listed the implications that were to be included in the conclusion section by using the *frame markers* "first… third…". In doing this, Momo made full use of her native language to brainstorm ideas, and to frame the ideas in a structured way, which could facilitate her subsequent drafting of the paper in English.

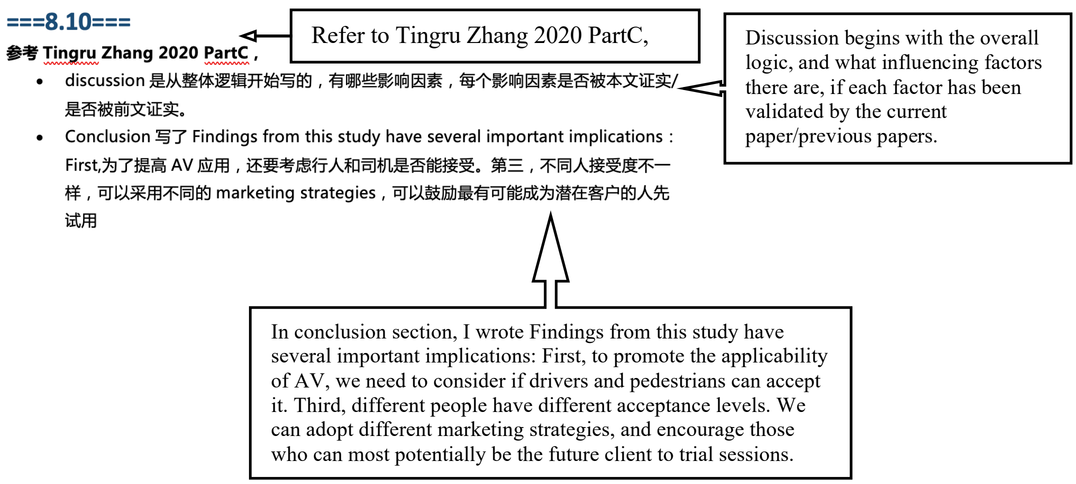

Figure 6. Extract from Momo's research log

Momo applied the brainstorming process using *frame markers* from her research log illustrated in Figure 6 in writing P3. In her P3, at the end of the paper, she wrote:

> <u>The results of this study have several significant ramifications. First,</u> to improve AV applications, the acceptability of vulnerable road users should be emphasized… <u>Second,</u> … This research guide AV shuttle operating agencies to identify user groups with positive initial opinion towards AV and the group who are inclined to change their mind properly and pinpoint the potential customers with the opportunities of AV ride experience. (P3)

Though we can see from the extract that wording was altered to some degree, the structure and content largely followed the notes she took in the research log about the implications of her study.

Though Chinese as her native language had assisted Momo in the initial phase of writing the papers to brainstorm ideas and structure the paper, that being an English as an L2 writer did bring her some confusions as she found it hard to distinguish the connotations of some *hedges* or *boosters*. For instance, in P3, she used expressions when reporting the results such as "suggest" and "indicate" which denoted hedging and uncertainty, as well as "show", a *booster*, which indicated certainty, as illustrated in the following three examples:

> This χ^2 value suggests that it is not significant that female participants have different parameter values…
>
> The results indicated that people's initial opinions toward the comfort of AV shuttles are significantly associated…
>
> The result shows we are over 95% confident in rejecting… (P3)

When asked what considerations she had made when using those words, Momo was surprised by my question, saying

> Momo: indicate, suggest and show, and hmm, don't they all mean the same thing?
>
> Ren: To you they mean the same right?
>
> Momo: Yes, they mean the same… I thought they can be used interchangeably, so I used all of them in my papers. (I2)

Then Momo explained that being a Chinese person whose first language was not English, she could not differentiate those words:

For example, like "indicate", which conveys uncertainty, to me whose first language is not English, maybe I can't realize the differences. It's good to know (the differences). (I2)

4.1.3.1.2 Being an anonymous reviewer

As depicted in Momo's profile in the methodology chapter, Momo had reviewed papers for journals. This experience of hers, which constituted her *autobiographical self*, had informed the identity construction in the way that it gave her a perspective of a reviewer, and endowed her with the understandings of what were expected of the writers. She then could construct an identity preferred by the reviewers. The influence of being an anonymous reviewer for journals on Momo's own academic writing, according to her, was "profound". She recollected her rationales for rejecting papers that she considered were not suitable for publication, which more or less included errors related to metadiscoursal usage such as *code gloss* and *endophoric markers*, as she said in the follow-up interview:

> If the paper I review has some most fundamental issues, such as diagrams/Figures/Tables containing unclear information, lacking annotations, mistakes in the numbering of formulas, initialisms/abbreviations not exactly represented, etc. My first reaction would be that it's a rejection. (FI)

In this quote, the presentation of diagrams/Figures/Tables and the numbering of formulas were related to the use of *endophoric markers*, which should be consistently used to locate corresponding information in a paper, for the sake of facilitating readers' processing of the knowledge. And the initialisms/abbreviations could be related to the use *code gloss*, as the information was provided in a concise form using those *code glosses*.

Given this, Momo mentioned that reviewing other writers' manuscripts prompted her to think about the presentation of the information in her papers in ways accepted by readers. She particularly was very cautious in making sure the multimodal information such as Tables and formulas could neatly and explicitly convey the knowledge she wanted to highlight; this had informed her conscious use of certain *endophoric markers*, as she stated,

> I think reviewing papers has a profound influence on my own writing. That is, I will spend a lot of time facilitating the image of the paper that is firstly conveyed to the reader. I will present diagrams/Figures/Tables nicely, and make sure the format is right. I will adjust it so that people are willing to reader my paper… (Obvious mistakes) bring bad impressions to reviewers, so I will consciously avoid that. (FI)

Momo brought up the importance of consistency of the multimodal information and their corresponding language describing the information:

> The diagrams/Figures/Tables should be consistent with the language. For example, if there're three main foci in a chart, I'd specify that in the paper, so that people can clearly understand what the chart is about. (FI)

Further, Momo articulated that as a reviewer, she valued the importance of specifying contributions of the papers in the introduction section. As she said,

> What I think reviewers would focus… since I also review other people's papers, I feel when I read their introductions, I'd focus on the contributions. First I'd focus on if it really is a contribution, then how innovative it is. Next I'd focus on if they can realize those claims, how well they were realized. So I think they (the reviewers) also review my paper the same way. (I1)

She mentioned the two ways she considered equally effective in doing so: using bullets and listing the contributions one by one in one or two paragraphs. This influenced her way of writing as an academic writer, as was demonstrated in her P1 and P3, where *endophoric markers* and *frame markers* were found to organize the contributions of her studies so that they were more approachable to reviewers. In P1, she used bullets to list the contributions (an *endophoric marker* was used), while in P3, an *endophoric marker* and a *frame marker* were employed:

> Motivated by these research gaps, <u>this paper presents an advanced control approach for roundabouts with three features:</u>
>
> - Improving the mobility and capacity of roundabouts;
> - Ready for real-world implementation;
> - Easy upgrade for partially connected and automated traffic. (P1)
>
> <u>The main contributions of this paper are two points.</u> <u>First,</u> we study how the quality of ride experience… <u>Second,</u> we consider people's willingness to share transportation… (P3)

Momo voiced her considerations for adopting these ways of textual representation:

> About the presentation of contributions, I feel either way is fine, because it’s not about which is longer, which is shorter. If you present the contributions in bullets, it means you have explained the gaps in the previous paragraphs. But discussing them in one or two paragraphs may be easier for readers to understand the logic of your arguments if they only read the contributions instead of the introduction section word by word. (FI)

4.1.3.2 *Possibilities for selfhood* and *discoursal self* and *self as author*

The *possibilities for selfhood* that provided a repertoire of identity in Momo's three papers entailed writing conventions from her disciplinary field, guidance from her community members, and academic courses she took, or lack thereof.

4.1.3.2.1 Writing conventions

Writing conventions in the engineering field shaped Momo's papers considerably. They mainly entailed the disciplinary writing conventions and journals' rules prescribed for writers to follow. The former influenced Momo's construction of herself as a disciplinary insider in the three papers through the use of multiple metadiscoursal markers. For example, offering an overview of the structure of the paper in the introduction section, according to Momo, was a commonly found feature in the papers in the field of engineering; hence she tended to keep this in her own writing. Accordingly, the last paragraph in the introduction section in Momo's three papers was mainly to offer an overview of the article through *frame markers* and *endophoric markers*, as in P1 in the example:

> <u>The remainder of the paper is organized as follows:</u> <u>Section 2</u> describes the problem considered in this study; <u>Section 3</u> describes the proposed controller and its solving process; <u>Section 4</u> conducts simulation evaluation and sensitivity analysis. <u>Finally,</u> conclusions and recommendations are delivered <u>in Section 5.</u> (P1)

Similarly, in the beginning of the conclusion section in P3, Momo recounted the main findings of her studies:

> <u>This section presents a summary</u> of all significant variables found at least in one of the initial opinions and opinion change models. <u>They are summarized in Table 9.</u>
>
> 1. Participants feel less comfortable about interacting with the AV shuttle…

2. Hard breaks deteriorate the attitudes about…

3. We can observe that participants with long commute times intend…

4. The seniors and females are more conservative about driving on the same route…

(P3)

In the following interview extract when I consulted her about presenting an overview of her paper, Momo stated that her reason for using the metadiscoursal markers was due to the writing convention in the engineering field:

Ren: Here, at the end of introduction you wrote a paragraph of "the remainder of this paper is as follows…". Is this your personal writing preference?

Momo: People in our field will write this way. (I1)

Additionally, in our first paper interview (I1), she mentioned the presentation of the information using *frame markers* and *endophoric markers* was due to the preferred writing style in the field:

Ren: Here you wrote the review of …, and the three points, which are the summary of the results. Was this required by the journals or was it your own writing style?

Momo: A lot of people (in our field) write this way. It's also much clearer. (I1)

Momo's employment of *directives* was influenced by the writing conventions in her field. However, the choices of the specific linguistic devices were determined by Momo. For example, in P3, toward the end of the paper, she suggested a few directions for future research as well as for policy makers. The sentences she wrote were as follows:

… the acceptability of vulnerable road users <u>should</u> be emphasized.

Policymakers <u>should</u> pay attention to the interactive experiences of vulnerable road users…

Transit agencies <u>should</u> focus on rigorous assessment and validation… (P3)

The recommendations were characterized by the overreliance on “should”, the *directive* indicating Momo’s dominant position as the writer of the paper. Momo explained that in her field, the content, instead of the language itself, of the suggestions for practical implementations was more important:

> In the end of the paper, the sentences are important. In our field, introduction is the most important, conclusion is the second. My purposes are to tell everybody what I find and how the findings are useful to guide government to do something. So I think that I use “should”, or how strong I sound is not important. Pointing out the direction, if it’s a correct direction, is more important. (I2)

On the other hand, Momo’s metadiscoursal choices that contributed to her identity construction were also shaped by journals’ rules. She mentioned that she “already had a targeted journal when she has not finished revisions on P1” (I1). The notable example that journal conventions affected her metadiscoursal choice was the use *self-mentions*. As reported in the previous sections, Momo exclusively used “we” to refer to the authors of the articles, which sometimes was inclusive of the readers. When consulted on the use of self-mentioned pronouns and when she would choose to use these pronouns, she remarked that articles from the top journals in her field altered her opinions of using *self-mentions*:

> When I did my undergrad, I remember vaguely I read from online that using “we” seemed not professional, so self-referenced pronouns should be avoided. But later on I saw articles published in top-tiered journals using “we”, so then I thought using it or not may not be that important. (I1)

4.1.3.2.2 Guidance from the community members

Momo's choices of metadiscoursal features were guided by various groups of community members that included her advisors, senior co-authors and journal/conference reviewers, and they shaped from various ways her construction of identities.

4.1.3.2.2.1 Advisors

Momo's advisors from both the master's program and the doctoral program shaped her identity construction mainly in terms of the use of *transitions,* which made her texts more logically coherent and reader-friendly. As pointed out in the methodology chapter, Momo's P1 was a collaboration with an advisor from her master's program, who helped her reshape the paper throughout the drafting and revision process. Momo spoke highly of her advisor's role in revising the paper "word by word"; she said, "If you see any differences between this version and the previous version, if you feel my language use and logic reasoning have improved, that's a credit to my advisor" (I1).

However, Momo remarked that in the current PhD program, since she was still a junior PhD student, she mainly worked with her advisor's previous advisees, whose role in shaping Momo's papers will be presented in the next sub-section. Momo stated her PhD advisor did help revise some of her reports and he focused overwhelmingly on the logic of her writing, and Momo's writing was criticized as lack of logic by both of her advisors in the master's and doctoral programs. As she said, "The most serious problem in our writing (here referring to another peer from Momo's cohort) is we can't write things in a logical way. We talk about this first, then that. Then we talk about another thing followed by something else" (I1). In the background interview, Momo also mentioned that she found logic and flow in academic writing to be extremely challenging, which she worked hard to improve but only ended up with dissatisfactory outcomes.

In our interview on the survey paper (I2), Momo articulated that under her PhD advisor's supervision, she now had made some improvement in regard to logic, "(I made some progress in paper writing.) Now I think I can write my papers in a logically sounder way than before". Momo shared the specific ways her advisor preferred in academic writing – framing the sentences in the sequence of ABBC. That is, "starting with a sentence, and the following sentence should include part of the content from the previous sentence." (I1). This could probably be illustrated in the following example: Momo first used an *endophoric marker* and referred to the literature reviewed in the texts above, and then used a *transition* marker, "however", to point out the limitations the previous studies had. This way, the logic could be made clear and reasonable, smooth for readers to follow:

> The aforementioned research about public attitudes focuses on the feelings about using the AV technique. However, this perspective overlooks the various road users who will be directly affected and forced to interact with AVs. (P3)

4.1.3.2.2.2 Senior co-authors.

Senior researchers from Momo's program played an immense role in Momo's drafting and revision of the three papers. In all of the three papers, Momo's seniors were co-authors. Two authors of her P2 and P3 were past advisees of her advisor and they had graduated from Momo's doctoral program. They influenced Momo's writing in the use of certain metadiscoursal markers. For example, in the feedback the senior co-authors offered on P3, one of them particularly pointed out the confusion caused by a term that was mentioned the first time in the results and discussion section of the paper - "This is not a terminology everybody knows. Define it properly". By offering definitions of a term that appeared the first time, the senior co-author's comment helped Momo present the information in a more professional way through *code gloss* in the subsequent drafts, the lack of which may bring confusion to readers, as indicated by senior's the comment.

Besides interactive metadiscourse, the seniors' revision held sway on Momo's expression of the degree of certainty. For instance, in P2, Momo used a phrase "The jury is still out", which was an informal *hedging* expression, but in P3, the revised version of P2, this phrase was removed. When asked about the reasons, she said,

> At first, I wanted to express "this is uncertain" in a different way, then I googled and I saw this phrase… Later I removed it because the co-authors said, "What the heck is this?" (I2)

Noticing that "The jury is still out" was an inappropriate expression to use in the research articles genre, Momo followed the co-authors' comments and deleted this phrase.

Not only did the senior co-authors focus on the appropriate usage of *hedges*, they also pointed out, on the contrary, that Momo sometimes appeared to be too assertive by using *boosters* "inappropriately". For example, the following sentence "it is definitely the vulnerable pedestrians who are more seriously hurt." was found in P2, but according to Momo, "the sentence was removed, because my co-authors said it's too much. The tone is too much. Similar meaning can be expressed without the assertive word, so we deleted the word at least" (P3). Further, in the conclusion section of Momo's P3, a categorical stance was noticeable in the sentence "only technically mature AVs should be permitted for public roads", which received the comment "This is too aggressive a statement" by the senior co-author. Using both *boosters* (only) and *directives* (should) did make the sentence take on a categorical tone, so the senior's feedback prompted Momo to revise the sentence in a less imposing way.

4.1.3.2.2.3 Manuscript reviewers.

As a first and second year PhD student, the *possibilities for selfhood* Momo gained from the community resources shaped her *discoursal self* and *authorial self* in multiple ways, and helped her juggle among different words and expressions to gradually construct an image of a competent writer.

Momo received manuscript reviewers' feedback on her papers, which she found "useful" (I2), as she said:

> The conference is the biggest annual conference in our field. I will not publish my paper at the conference. It also will not affect me submitting it for journal publication later on… And they (reviewers) gave me a lot of comments. One reviewer was an expert, and gave me a lot of valuable feedback. (I2)

On her P2, which was submitted to an annual conference in her field, the reviewers specifically highlighted the important role of the *frame markers* that Momo used to list the important findings: "I appreciated the summary bullets and summary table 10 in the end of results." Moreover, another reviewer also pointed out the problems that the *endophoric markers* in Momo's paper may have: "The titles of Figure 5 and Figure 6 seem to be the same. Figure 6 may be used to show people's perception of comfort as a pedestrian before and after the test ride". Momo revised accordingly, and thus, it was more convenient for readers to connect the linguistic descriptions and the Figures. The reviewers' comments facilitated Momo's construction of a professional writer identity that deploy metadiscoursal resources such as through *frame markers* and *endophoric markers* to better connect her writing and facilitate readers' understanding.

4.1.3.2.3 Academic courses

Despite the fact that Momo had academic writing experiences since college, she did not have specific courses designated for academic writing. As described in Momo's profile in the methodology chapter, during her undergraduate phase, the academic courses offered by Momo's program were related to citation format, avoidance of plagiarism, correspondence writing, and so on. In her master's program, she did not have English academic writing courses. Realizing that academic discourse was a completely different and new set of language system, Momo felt frustrated about the challenges she was always faced with in academic writing in English, especially the logical flow of ideas, which could be reflected in the use of metadiscourse markers such as *transitions*. As she said, "I feel there's no way to deal with that… Also there's no course teaching it (how to write logically)." (BI)

However, the only academic writing training that benefited Momo was the TOEFL and GRE writing courses she took off campus, as she said,

> I feel I didn't get enough training in this regard (writing logically in academic writing). You know the actual academic writing experience I had or the stuff I learned useful for academic writing was probably only from the TOEFL and GRE writing courses… The biggest restriction for me is logic. To me this is a completely new area. It's a different system from the English we learned from elementary school. (BI)

Up till data collection, Momo did not take any academic writing courses in her PhD program. Though she was required to submit writing assignments for the courses she took in the program, the focus was more on technical aspects such as algorithm, and "they don't specifically focus on language" (I1).

In a nutshell, given the difficulties Momo encountered in academic writing, which may have constrained her identity construction as a scholarly writer, the lack of corresponding academic training that could have been available in academic writing courses constituted a factor which negatively affected Momo's identity construction in her papers, particularly in regard to the connection of ideas logically through linguistic devices, such as *transitions*.

4.1.3.3 Agency and *discoursal self* and *self as author*

Besides the aforementioned sociocultural and academic influences on Momo's identity construction in the three papers, Momo also actively chose from the linguistic repertoire available to her to construct the identities she deemed acceptable. This was demonstrated in her consideration of the nature of information presented in the text, and avoiding repetition, which are reported below.

4.1.3.3.1 Nature of content

Momo argued that whether the information was a fact generated from programming and calculation or her own opinions was the main reason to express different degrees of certainty in the paper. In P1, for example, the sentence "its fast calculation guarantee that it is ready for real world implementation" was found. When asked why she used a *booster* (guarantee) to express assertiveness in this sentence, Momo said:

> If our control of CAV reaches 0.1 second, generally if the time calculated is lower than 0.1 second, I'd think it can be used in reality. The precise control is 0.1 second. Before

> this I have already proved that the time I calculated is way lower than this, so I'm certain when I wrote the sentence… I have statistics to support this. (I1)

She also mentioned that "if it cannot be proved by experiment or the data from previous research, I would not use this kind of words (*boosters*)." (I2)

Different from the certainty that can be drawn from objective experimentation, some *hedges* were used purposely by Momo due to her subjective judgements. In P1, Momo wrote sentences entailing *hedges* such as "there is a chance…" and "it is suggested to…" as shown in the following examples,

> <u>There is a chance that</u> an earlier special temporal resources is not significant enough for a higher priority vehicle…
>
> <u>It is suggested to</u> put in larger objects before the smaller ones. (P1)

Specifically explaining the second example, Momo voiced the reasons for using the *hedge* and being cautious here:

> Here it depends on specific situations… I was explaining why left turning cars should be given priority, because I think if they were, it will improve the overall efficiency. There exists this possibility, and I can't be 100% sure prioritizing left-turning cars will improve that, so that's my explanation. (I1)

Similarly, in P2 and P3, Momo used *hedges* to offer interpretations of the research results; as the following examples show, "possible" and "might" were employed:

> <u>There are two possible reasons for</u> this result. (P3)
>
> <u>This might be because</u> they had preconceived notions about how… (P2)

That Momo did not have experimental data to prove the certainty of the statements was the reason for using *hedges.* As she said,

> Here it should be what I said when explaining the results. For example, if I only have the results that people with higher income tend to have negative attitude, I can only conjecture the reasons for that. If I want to know the exact reasons, I have to do new experiments. (I2)

Further, Momo used *directives* strategically depending on the content she wanted to highlight, which she perceived may be better than previous research. She used the following sentence in P3,

> ... ε_n is the disturbance term. <u>Note that</u> only one participant chooses "very uncomfortable". (P3)

She explained the usage this way, “Usually if I use ‘note that’, it should be about my methods. If mine are different from that of previous research, and mine is better than others’, I would possibly use ‘note that’.” (I2) This demonstrated that Momo chose to use *directives* in her papers sometimes for the purpose of highlighting and differentiating the information in her papers.

4.1.3.3.2 Avoiding repetition

Momo actively chose from the available linguistic devices to promote the diversity of linguistic choices in her papers. This was conspicuous in her use of *self-mentions*. As was pointed out in the methodology section, Momo’s P3 was a revised version of P2, and would be submitted for publication soon. She mentioned that in the revised version, she either removed “we” from some sentences or substituted “we” with “this project”, “this work” or “this paper”:

> Momo: The P3, I revised it and I deleted “we” in some places. In other occasions, I feel it’s probably better to use this paper, this work, I changed it to “this paper”. I just didn’t use “we” in the entire paper…. Yes, I revised “we” in the paper.
>
> Ren: So do you think using “this paper” or “this study” is better than using “we”?

> Momo: I think it is not about if one word is superior than another. It's just, don't use the same word too many times. (I2)

However, even though she consciously chose to remove some self-referenced pronouns, in P3, there were still more cases of *self-mentions* than in the other two papers. Momo exerted agency to maneuver the degree of the conspicuity of herself and other authors in the paper. The juggling between appropriate and available expressions showed how her agency played a role, which strategically shaped her construction of *discoursal self* and *authorial self* in the papers.

4.2 Nancy

In this section, I report findings related to Nancy's identity construction in her three papers, starting with the distribution of metadiscourse markers, followed by patterns of metadiscourse use in identity construction in the papers, and the factors influencing her identity construction in the three papers.

4.2.1 Metadiscourse usage in identity construction in Nancy's papers

In this section, I present how Nancy's identities were constructed through the use of metadiscourse, first through the overall distribution of metadiscourse in Nancy's papers. Then I report the distribution of each category of metadiscourse starting with interactive metadiscourse, followed by stance and engagement markers under interactional metadiscourse.

4.2.1.1 Overall distribution of metadiscourse markers in Nancy's papers

As presented in Table 13, the overall pattern of metadiscourse use in Nancy's papers was consistent, with the overwhelmingly used category being interactive markers, followed by stance markers, and engagement markers. To look at the total number, altogether there were 1,754 cases of metadiscourse markers in Nancy's three papers, with interactive metadiscourse being the most frequently used category, covering 72.5% of the total metadiscourse markers, 29.1 cases per 1,000 words. The second most frequently used was stance markers, constituting 25.9% of all the metadiscourse markers, which was 10.4 times after being normalized per 1,000 words. The least used was engagement markers, covering 1.6% of all the metadiscourse markers, only 0.7 times every 1,000 words.

The frequency of each of the metadiscourse categories showed largely similar patterns across Nancy's three papers: interactive metadiscourse was used overwhelmingly, whereas engagement markers were used negligibly. Nancy's P1, which was a research article submitted for journal publication, was found to have 202 cases of metadiscourse markers. Among them, interactive metadiscourse occurred 146 times, covering 72.3% of the total metadiscourse markers. Stance markers appeared 53 times, covering 26.2% of the total metadiscourse markers. Engagement markers were only used 3 times.

The second piece of paper of Nancy's (P2) was written for the government reporting the progress of a project undertaken in the research center where Nancy was working. In P2, there were 610 cases of metadiscourse markers, 32.8 times per 1,000 words. Interactive metadiscourse was used 427 times, covering 70% of the total metadiscourse markers. 175 cases of stance markers were found in P2, constituting 28.7% of all the metadiscourse markers. Engagement markers, however, were found to appear 8 times, only 0.4 times per 1,000 words in P2.

Table 13. Overall distribution of metadiscourse markers in Nancy's papers.

	P1		P2		P3		Total	
Category	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words
Interactive metadiscourse	146 (72.3%)	23.6	427 (70%)	22.9	698 (74.1%)	37.1	1,271 (72.5%)	29.1
Interactional metadiscourse (stance)	53 (26.2%)	8.6	175 (28.7%)	9.4	227 (24.1%)	12.1	455 (25.9%)	10.4
Interactional metadiscourse (engagement)	3 (1.5%)	0.5	8 (1.3%)	0.4	17 (1.8%)	0.9	28 (1.6%)	0.7
Overall total	202 (100%)	32.7	610 (100%)	32.8	942 (100%)	50.1	1,754 (100%)	40.2

Lastly, Nancy's P3, her dissertation, was found to have 942 cases of metadiscourse. Among them, 74.1% were interactive metadiscourse, with 698 items being found. Stance markers covered 24.1% of all the metadiscourse markers, which was 227 cases in total. Engagement markers, however, were found to be used 17 times, covering 1.8% of the total metadiscourse markers.

4.2.1.2 Interactive metadiscourse and identity construction in Nancy's papers

Table 14 displays the interactive metadiscourse usage in Nancy's three papers. As has been presented in the previous section, this category of metadiscourse was used overwhelmingly across the three papers – they all covered more than 70% of the metadiscourse markers. As an important linguistic device to signal the explicit relations among sentences, arguments and propositions, the extensive occurrence of this metadiscourse category showed Nancy's efforts (or habit) to link sentences together in a coherent and logical manner, demonstrating her identity as an academic writer who was able to present the internal relations of her papers in a coherent manner.

Surprisingly, in Nancy's three papers, the most frequently used interactive metadiscourse was *code gloss*, covering 34.2% in P1, 32.8% in P2, and 28% in P3 respectively. The high frequency showcased that Nancy was extremely cautious about making clarifications in regard to certain phenomena or interpretations, so that readers would not misinterpret or misapprehend her intended meaning. For example,

> Two normally-distributed random variables (segment length and access density) were identified after 300 Halton draws in both the nighttime model and the daylight model. (P1)
>
> Major arterials and minor arterials account for the majority of MCCS crashes (45.6% and 28.9%, respectively). (P2)
>
> The sample size of MCCS data was too small (up to 23 observations for each crash type) to apply regression and hypothesis testing; (P2)
>
> As a variety of contributing factors (e.g., tire marks, debris, roadside objects, traffic volume) were creatively investigated and documented… (P2)
>
> Compared to other shoulder types (i.e., raised and lawn), paved shoulders provide additional space for motorcycles and vehicles to take non-typical actions such as overpassing, speeding up, etc. (P3)

As shown in the examples above, Nancy had a preference to offer additional information in parentheses for further elaboration. This was helpful in achieving the interpersonal purposes in academic papers, which could make sure the writer's intended meanings were accurately delivered to the audience. When addressing a potentially diverse group of readership, Nancy's use of *code gloss* extensively demonstrated her "audience-sensitivity" (Hyland, 2007, p. 266) and her identity as an authoritative writer who knew well the content knowledge, and who had taken readers' needs into consideration so that the meanings could reach the readers unambiguously.

The next sub-category of interactive metadiscourse that was frequently used across Nancy's three papers was *evidentials*. *Evidentials* were densely found in the introduction sections in Nancy's papers, where numerous sources were cited, as shown in the following examples:

> …motorcycle crash studies are often constrained by the limited motorcycle-specific information in police crash reports (*1–3*). (P1)
>
> Most studies considered street lighting as a binary factor (presence or not, or before and after improvement) at intersections (Bruneau and Morin 2005; Bullough et al. 2013; Donnell et al. 2010; Elvik 1995; Gbologah 2015; Isebrands et al. 2010; Kim and Washington 2006; Preston and Schoenecker 1999), roadway segments (Anarkooli and Hadji Hosseinlou 2016; Wanvik 2009; Yu et al. 2015; Zhang et al. 2012), or both (Monsere and Fischer 2008; Sullivan and Flannagan 2002). (P2)
>
> Speeding is a predominant factor that tends to increase the likelihood and severity of motorcycle crashes (Chang and Yeh, 2006; Chen et al., 2014; Haque et al., 2010b; HARNEN et al., 2003; Jung et al., 2013; Quddus et al., 2001; Rifaat et al., 2012b; Savolainen and Mannering, 2007; Schneider and Savolainen, 2012; Shaheed and Gkritza, 2014b; Shankar and Mannering, 1996). (P3)

Citations are a distinctive feature in academic written discourse. In Nancy's papers, a list of sources was presented, which showed the extensive knowledge base she had accumulated over time. Further, integrating a web of intellectual resources in the papers constructed Nancy's identity as a knowledgeable researcher and student (considering the different purposes of the papers), who had been well-informed of the field knowledge, and who could make reasonable claims based on her understanding and evaluations of the disciplinary knowledge.

Table 14. Interactive metadiscourse in Nancy's papers

	P1		P2		P3		Total	
Category	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words
Transitions	24 (11.9%)	3.9	70 (11.5%)	3.8	162 (17.2%)	8.6	256 (14.6%)	5.9
Frame markers	3 (1.5%)	0.5	10 (1.6%)	0.5	21 (2.2%)	1.1	34 (2%)	0.8
Endophoric markers	12 (5.9%)	1.9	102 (16.7%)	5.5	58 (6.2%)	3.1	172 (9.8%)	3.9
Evidentials	38 (18.8%)	6.1	45 (7.4%)	2.4	193 (20.5%)	10.3	276 (15.7%)	6.3
Code gloss	69 (34.2%)	11.2	200 (32.8%)	10.7	264 (28%)	14	533 (30.4%)	12.2
Interactive (Total)	146 (72.3%)	23.6	427 (70%)	22.9	698 (74.1%)	37.1	1,271 (72.5%)	29.1

Thirdly, another interactive metadiscourse that was rather frequently used was *transitions*, covering 11.9%, 11.5% and 17.2% of metadiscourse use in Nancy's three papers respectively. *Transitions* served as the explicit signals of the relations among sentences. Nancy employed *transitions* rather densely to link the texts together. Consider the extracts in the following examples:

> … driver vision is weakened during the adjustment period and, consequently, may decrease visibility and increase driving risk at night. The standard deviation of horizontal

illuminance on a whole segment, however, is not necessary to capture the true lighting pattern related to nighttime safety… (P1)

Although studying motorcycle crashes has become increasingly common, few studies have considered causal factors or attempted to build an in-depth understanding of those crashes. MCCS data could help to fill this research gap, as more than 2,000 motorcycle crash features are included… Additionally, MCCS built detailed data-collection forms for case-control studies and training materials for future large-scale studies. (P2)

The RBP-R model shows that riders carrying insurance tend to take less improper pre-crash actions; correspondingly, they are about 0.4% less likely to be killed in a crash on curves. (P3)

In these sentences, Nancy used *transitions* of various kinds rather extensively, including conjunctions (and, however, as) and adverbials (consequently, correspondingly), which could reflect her metadiscoursal considerations of readers' ability to recover the intended connections (Hyland, 2018). Through the use of *transitions*, Nancy was constructed as an author who could demonstrate her field knowledge in a coherent way, and through Nancy's making explicit the "surface logicality" (Hyland, 2018, p. 129), readers did not have to code the relations between propositions based on their understandings of the knowledge in Nancy's papers, where it was the author/s themselves who were responsible for the meaning making process.

Endophoric markers was not a highly frequently used interactive metadiscourse category in Nancy's three papers, constituting 5.9%, 16.7% and 6.2% of the total metadiscourse markers in P1, P2 and P3 respectively. As explained earlier, *Endorphoric markers* are a range of linguistic markers that could also function to glue the texts together, making referring to other parts of the text explicit and straightforward, as shown in the following examples:

> The illuminance data were imported into a GIS layer and spatially matched to each site in ArcGIS, <u>as shown in Appendices</u> (Figure A1). (P1)
>
> <u>A summary of objectives, data processing procedures, methodologies, and major findings for the five studies is presented in **Table 1**.</u> (P2)
>
> The sampling procedure is shown <u>in Figure 4.3</u>. (P3)

In the hard sciences field, "endophorics [are] central to scientific argument" (Hyland, 2018, p. 191), and an extensive amount of Tables, Figures, and formulas is used to present the information in a more organized way (Hyland & Tse, 2004), which can help readers to visualize information more efficiently. *Endophoric markers* were commonly found in Nancy's papers to direct readers' attention to locate the corresponding materials. The use of this metadiscourse marker demonstrated Nancy's identity as a writer who was cognizant of readers' processing needs and the disciplinary conventions, and who in the meantime also made sure that readers had direct access to the materials located elsewhere in the text (Hyland, 2018).

Lastly, *frame markers* were the least frequently used interactive metadiscourse, appearing 3 times, 10 times, and 21 times in Nancy's three papers, covering 1.5%, 1.6% and 2.2% respectively of the total metadiscourse markers. As has been reported in Momo's results, frame markers encompass four sub-categories, including announcing goals, shifting topics, labeling stages, and sequencing. Similar to in Momo's papers, announcing goals, labeling stages, and sequencing were found in Nancy's papers. The following two extracts exemplified frame markers stating the goals of Nancy's papers:

> <u>The primary objective of this study</u> was to investigate the effects of photometric characteristics (horizontal illuminance pattern) of street lighting… (P1)
>
> Based on the research gaps revealed in the introduction, this dissertation <u>aims to</u> answer

the following research questions. (P3)

Next, *frame markers* were found to function to label stages in Nancy's papers, as shown in the following examples:

> In summary, it is clear that the MCCS collected many unique variables that provide considerable insight regarding contributing and causal factors for motorcycle crashes. (P2)
>
> Overall illuminance standard deviation may represent the difference of lighting patterns… (P1)

Sequencers were also found in Nancy's papers to break down a notion into more specific points. For example, in the following sentences, listing what the challenges entailed one by one could potentially make the cognitively demanding notion more accessible to readers:

> There are several challenges in modeling the relationship between street lighting uniformity and nighttime crash risk. First, big scale lighting data in… Second, nighttime crash observations are very rare… Third, all studies of lighting photometric measures were based on… Fourth, to increase the sample size of nighttime crashes on… (P3)

The use of *frame markers* indicated Nancy's efforts in structuring the unfolding discourse. This helped construct Nancy's identity as a skillful academic writer who could provide explicit signposts where appropriate to lead readers throughout the discussion.

4.2.1.3 Stance markers and identity construction in Nancy's papers

In this section, I present Nancy's use of stance markers for identity construction. The employment of stance markers is displayed in Table 15. As shown, the four subcategories of stance markers were used in all three papers, except that *self-mentions* were not found in Nancy's P2, the report for government.

The most frequently used stance marker category was *hedges*, which constituted 18.3%, 21% and 13.1% of the metadiscourse markers across Nancy's three papers. As can be seen in the following extracts, Nancy employed various kinds of linguistic resources to express the uncertainty and speculations:

> The regression method based on cross-sectional data may not fully address the interferences of cofounders. Matched pairs studies (e.g., matched case-control study) may be applied to address this issue. (P1)
>
> It is possible that this finding is related to an increased rate of distracted driving, a prospect that warrants further investigation. (P2)
>
> …and aggressive curve negotiating behaviors are believed to be primary causes contributing to severe injuries… (P3)
>
> To the best of our knowledge, no work has been done on using collected structured data and satellite images to predict road safety. (P3)

The examples were typical of Nancy's using of *hedges* in her three papers. It is clear that the propositions were based on Nancy's own point of view, so she was cautious about not making sweeping claims that may sabotage the credibility of the propositions as well as of herself as an author. Adopting a reduced and attempting tone applied to the propositions showed that Nancy potentially admitted the other possible and reasonable interpretations. Through the use of *hedges* to downplay the assertiveness of her own opinions, Nancy was constructed as a skillful writer who could present claims appropriately in light of the extant disciplinary knowledge, and make personal judgements. However, it needs to be pointed out that Nancy demonstrated the cautious identity through use of “may”, this was not purposeful. This will further be discussed in the discussion section.

The next frequently used stance marker category was *boosters*, covering 2.5%, 4.9% and 7.6% of metadiscoursal markers used in the three papers respectively. Different from the opinions that were often associated with the use of *hedges*, *boosters* were found more often than not related to statements that could be supported by some evidence, as in the following examples:

> Particularly, this study aimed to (1) address the unobserved heterogeneity issue using random parameter models… (P1)
>
> … it is strikingly clear that alcohol/ drug/alcohol and drug-impaired crashes accounted for a significantly larger proportion of LTAP motorcycle fatalities (27.9%)… (P2)
>
> It is obvious that number of safe spots is more than number of crash spots, while the true distribution of crash spots in the roadway network is still unknown. (P3)
>
> These studies demonstrated the use of MCCS data on various topics through descriptive analysis or statistical modeling… (P2)

As indicated in the examples, Nancy drew on linguistic devices such as adverbials (strikingly, particularly) and emphatics (obvious, demonstrate) to strengthen the claims and emphasize the truth of a statement. The amplifying tone made the writer sound confident and assured when making the claims. It is quite clear that the propositions in the above examples were generated based on certain objectivity (e.g., experimentation), where using *boosters* could enhance academic persuasion, and constructed the author’s image as a writer who could project confident voices where appropriate.

Table 15. Interactional metadiscourse in Nancy's papers – stance

	P1		P2		P3		Total	
Category	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words
Hedges	37 (18.3%)	6	128 (21%)	6.9	123 (13.1%)	6.5	288 (16.4%)	6.6
Boosters	5 (2.5%)	0.8	30 (4.9%)	1.6	72 (7.6%)	3.8	107 (6.1%)	2.4
Self-mentions	2 (1%)	0.3	0 (0%)	0	14 (1.5%)	0.8	16 (0.9%)	0.4
Attitude markers	9 (4.4%)	1.5	17 (2.8%)	0.9	18 (1.9%)	1	44 (2.5%)	1
Interactional - stance (Total)	53 (26.2%)	8.6	175 (28.7%)	9.4	227 (24.1%)	12.1	455 (25.9%)	10.4

Attitude markers were a subcategory of stance markers that was used less frequently than *hedges* and *boosters*, covering 4.4%, 2.8% and 1.9% of the metadiscourse markers respectively in Nancy's three papers. Upon further looking at the data, Nancy preferred to use adjectives to express her attitudes toward a research finding or a statement, as in:

> It is however contradicting to the intuition that the street light pattern should not have a relationship with daylight crash occurrence. (P1)
>
> It is not surprising that younger riders have a larger share of LTAP crashes, a fact that might be related to younger rider inexperience, lack of professional training, and careless riding. (P2)
>
> Observational cross-sectional studies without any control on confounding factors may lead to counterintuitive conclusions such as high lighting level associated with high nighttime crash frequency (Janoff et al., 1978; Wei et al., 2016)… (P3)

It is interesting to find that riders are more likely take improper actions in daylight compared to poor lighting conditions. (P3)

In the extracts above, Nancy not only presented the findings based on experimentation but also inserted her personal affective judgments toward those findings. Integrating affective attitude could personalize the academic discourse and make the academic texts more reader-friendly, as, according to Hyland (2018), this kind of expressions implicitly assumed that readers experience the discourse the same way as writers did. Using *attitude markers* in the papers positioned Nancy's persona as an agentive writer who not only could demonstrate professional expertise through careful data-supported research, but was also able to take one step further and evaluate those findings.

Self-mentions were the least frequently used stance marker in Nancy's three papers, which were not found in P2, and this sub-category constituted 1% and 1.5% of the total metadiscourse markers in P1 and P3 respectively. Nancy exclusively used "we" and the corresponding adjective pronouns to refer to herself (and the co-authors) in the three papers, as shown in:

Taking the logarithm for both sides in Eq.6, we have $LN(CMF_{A \to B})$ (P1)

Chapter 2 investigated the safety effects of roadway illuminance... We also developed reliable CMFs based on the econometric modeling results. (P3)

Our image-based model reaches the accuracy of 78%, similar to Najjar's study (Najjar et al., 2017). (P3)

As displayed in the examples, when referring to the experiment procedures, Nancy tended to refer to the authors themselves explicitly. This could help establish Nancy's (and the co-authors') active roles in undertaking the research studies, and constructed Nancy's identity as the agent who implemented the experiment or who calculated the algorithms needed in the studies.

4.2.1.4 Engagement markers and identity construction in Nancy's papers

In this section, I report findings related to engagement markers that constructed Nancy's identities in the three papers. In general, engagement markers were used minimally in all of Nancy's papers, covering 1.5%, 1.3% and 1.8% respectively of all metadiscoursal markers used, roughly at the same frequency.

Among the engagement markers, the two sub-categories that were used in all three papers were *directives* and *appeals to shared knowledge*, with *directives* being the most frequently used sub-category. As discussed in Momo's case, *directives* entailed *textual acts, physical acts*, and *cognitive acts*. In Nancy's papers, it was found that *physical acts* and *cognitive acts* were used to instruct readers. *Physical acts* were found primarily toward the end of the texts to point out how the research results could inform the implementation of the research findings in the real world or how future research should be conducted to extend the research agenda, as in the following examples:

> New uniformity measures <u>are needed to</u> capture the true interlacing pattern along travel routes… Further study <u>is needed to</u> address the endogeneity caused by selectivity bias in a sample. (P1)
>
> Researchers and engineers <u>need to</u> apply more efforts to reduce motorcycle crash risk for the two days. (P1)

> It is suggested that in XX[1], riders with USDOT-compliant helmets (with partial coverage) should be compared with riders with other types of helmets (*1*). (P2)

Using *directives* fulfilling the *physical acts* function showed that Nancy was constructed as an authoritative writer who could generate suggestions and recommendations to direct readers to some future actions based on her research findings.

Cognitive acts, on the other hand, were found in Nancy's papers to a lesser extent. In the following example from P3, Nancy used a parallel of *directives* to explain the constituents of a formula. In doing so, Nancy instructed readers to see things from a perspective determined by herself (and the co-authors), which performed the *cognitive acts* functions.

> Let i ($i = 1, 2, \cdots, I$) be an index to represent the matched case-control stratum… Let j ($j = 0, 1, 2, \cdots, J$) be an index to represent the observation record within each stratum. Let k ($k = 0, 1, 2, \cdots, K$) be an index to represent the unmatched explanatory variable xk. (P3)

Through the use of *directives* which performed *physical acts* and *cognitive acts*, Nancy was constructed as an author who could directly engage readers in the discourse, and bring them into the process of reasoning.

Another type of engagement marker that was found across the three papers was *appeals to shared knowledge*, which appeared once in P1 and P2, twice in P3, as in the following examples:

> This means that some comparisons should be treated with caution, as rider age is widely known to be causally associated with motorcycle type and activity. (P2)

[1] This was originally a word referring to the state in which Nancy's research was conducted.

This conclusion is consistent with two previous studies (Wang et al., 2017; Zhao et al., 2015) and the common sense that drivers need more time to adapt changes of high-to-low or low-to-high lighting patterns… (P3)

Nancy referred to the common knowledge shared by the readers and herself, assuming that readers had no difficulty understanding the propositions and that readers would agree with the statements. By doing this, Nancy constructed herself as a writer sharing the same disciplinary background knowledge with the readers, and signaled her identity as a disciplinary insider.

Table 16. Interactional metadiscourse in Nancy's papers – engagement

	P1		P2		P3		Total	
Category	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words
Reader pronouns	0 (0%)	0	2 (0.3%)	0.1	0 (0%)	0	2 (0.11%)	0.04
Directives	2 (1%)	0.3	5 (0.8%)	0.25	9 (1%)	0.5	16 (0.91%)	0.44
Questions	0 (0%)	0	0 (0%)	0	5 (0.5%)	0.25	5 (0.29%)	0.11
Appeals to shared knowledge	1 (0.5%)	0.2	1 (0.2%)	0.05	2 (0.2%)	0.1	4 (0.23%)	0.09
Personal asides	0 (0%)	0	0 (0%)	0	1 (0.1%)	0.05	1 (0.06%)	0.02
Interactional – Engagement (Total)	3 (1.5%)	0.5	8 (1.3%)	0.4	17 (1.8%)	0.9	28 (1.6%)	0.7

The rest of the engagement markers only appeared in one of the papers. As the most conspicuous signal showing writers' concern of the audience (Hyland & Jiang, 2016), r*eader pronouns* were found twice in P2, the government report, as in:

For example, augmented data collection and analysis will improve our understanding of various safety-enhanced riding behaviors… (P2)

> As such, MCCS data are expected to be helpful for expanding our understanding of motorcycle crash causation… (P2)

Through using “our”, Nancy included the readers in the text and identified with the readers, who were also the sponsors of the research report. She aligned herself (and other co-authors) with the readers and the two parties co-constructed an intellectual journey. By using “our”, she indicated that the readers agreed on the reasoning she made in the report and that they jointly arrived at the understanding of the issue under investigation. The use of *reader pronouns* in the government report constructed Nancy’s identity as an authoritative writer who born reader awareness and wrote in a reader-friendly way which could reach a more diverse readership. As Hyland and Jiang (2016) noted, in recent years, writers from the engineering field have increased their use of *reader pronouns* in research articles, which probably has to do with the pressure writers from engineering disciplines have to reach out to the potential readers outside academia to address the needs of the funders in the commercial world.

Questions, however, were only used in P3, and five cases were found in this paper, all of which appeared in raising research questions, as the examples below showed:

> Based on the research gaps revealed in the introduction, this dissertation aims to answer the following research questions.
>
> 1) What are the safety effects of …? How would roadway… affect the crash risks on a roadway segment?
>
> 2) What are the factors contributing to…? How could these factors affect the motorcycle fatalities on horizontal curves?
>
> 3) What is the appropriate method to recognize crash high-risk zone…? (P3)

Questions arouse readers' interest and invite readers to participate in the knowledge exploration with the writer. In Nancy's P1 and P2, *questions* were not used. In P3, they were raised in a separate section in the Introduction chapter. In P1 and P2, the purposes of research were presented in declarative sentences (for instance: "The primary objective of this study was to investigate…" in P1, and "this study will attempt to…" in P2). Though Nancy did not specifically talked about her reasons for using *questions* in P3, explicitly presenting research questions was a useful strategy in such a genre of extended length, which constructed her as a competent dissertation writer who probably subconsciously addressed readers' needs and invited them to her line of argument.

Personal asides were thought to be a feature belonging to more discursive field (Hyland & Jiang, 2016), and they only were used once in Nancy's P3, as shown in the following extract:

> Motorcyclists' safety compensation behaviors (they tend to take unsafe actions when they feel safe, such as in a good lighting condition) may cause this effect. (P3)

Nancy inserted the complementary remarks to the original statement to further elaborate on what the behaviors entailed. Rarely being used, the singular instance contributed to Nancy's identity as a writer who tended to clarify information for her readers.

4.2.2 Patterns of identity construction through metadiscourse use in Nancy's three papers

4.2.2.1 Overall patterns of metadiscourse usage in Nancy's papers

As displayed in Table 13, overall an overwhelming amount of metadiscourse in Nancy's papers was interactive metadiscourse, covering 72.5% in total metadiscourse use. The frequency across the three papers did not differ greatly (72.3%, 70%, and 74.1%). However, in P3, the dissertation, it seems significantly more interactive metadiscourse was found when normalized per 1,000 words.

Stance markers were used significantly less frequently than interactive metadiscourse, covering 25.9% in total metadiscourse use. When broken down, the frequency of stance markers was 26.2%, 28.7% and 24.1% in P1, P2, and P3 respectively, and it was in P3 that stance markers were used the most when normalized per 1,000 words. Engagement markers, on the other hand, were used minimally in all three papers, with 1.6% of the total metadiscourse items being engagement markers. Similarly, P3 had the most cases of engagement markers after normalization.

Taken together, each branch of metadiscourse markers was found the most frequently used in P3 after normalization, whereas in P1 and P2, the differences were noticeable, but it seems the differences were not remarkable. This probably showcased that when Nancy wrote P3, which marked the end of her PhD journey, she had learned more skills in academic writing that could assist her in making metadiscoursal choices to write internally coherent and reader-friendly academic papers. This tendency also demonstrated that she gradually was constructed as a competent and professional academic writer. The patterns of the sub-categories of metadiscourse used in Nancy's three papers are reported next.

4.2.2.2 Patterns of interactive metadiscourse in identity construction in Nancy's papers

As shown in Table 14, which displays the total interactive metadiscourse markers, *code gloss* was the most frequently used category, covering 30.4% of the total metadiscourse markers. The second mostly used interactive metadiscourse marker was *evidentials*, followed by *transitions*, and *endophoric markers*. *Frame markers* were found the least frequently; only 2% of the total metadiscourse markers belonged to this category.

Notable patterns in interactive metadiscourse use in Nancy's three papers can be found in *evidentials* and *endophoric markers*. In terms of *evidentials*, it was found that in P2, the report, it was used significantly less than the other two types of papers. In P2, 2.4 times per 1,000 words were found, while in P1, 6.1 times, and in P3, 10.3 times. *Endophoric markers* were used more frequently in P2 (5.5 time per 1,000 words) than in P1 (1.9 times) and P3 (3.1 times). Further, as the percentages of frequency show, *endophoric markers* constituted a major category of metadiscourse markers in P2 (16.7%), compared to P1 (5.9%) and P3 (6.2%).

The differences show that Nancy constructed herself as an author differently in writing for different purposes. In P2, where less *evidentials* were used, Nancy was constructed as the researcher and writer who knew well of the knowledge most relevant to the topic under investigation, where in other two papers, she reviewed more literature, which demonstrated her identity as a scholar who could pool disciplinary knowledge extensively to build a solid foundation for her own papers. The high frequency of *endophoric markers* in P2, however, could probably construct Nancy as an effective presenter of information who skillfully referred to other parts of a text in order to facilitate readers' understanding. This rhetorical requirement probably was less prominent in the other two genres.

Next, *code gloss* was the most frequently used metadiscourse in Nancy's three papers (11.2, 10.7 and 14 times after being normalized per 1,000 words in P1, P2, and P3 respectively). *Frame markers* were found to be used at the same frequency in P1 and P2 (both 0.5 times after normalization), while in P3, more cases were found, with 1.1 times per 1,000 words. Similarly, *transitions*, in P1 and P2, were used at a rather similar frequency (3.9 times vs. 3.8 times after normalization), but more (8.6 cases) were found in P3, the dissertation. The widespread employment of *code gloss* indicated that Nancy's identity as a writer who tended to rephrase propositional content to avoid ambiguity and facilitate readers' processing of the information, despite different genres. The use of *frame markers* and *transitions* was the most frequent in P3, which showed Nancy's tendency of gradually demonstrating herself over the years as a writer who was conscious of the internal relations among propositions, and tried to help readers to build links within her papers and make the connections easy to follow. This may also have to do with genre difference in that dissertation is a much longer piece of academic writing, which required the writer to employ *frame markers* and *transitions* to specify the textual relations explicitly.

In a nutshell, Nancy's P3 exhibited the most frequent use of most of the interactive metadiscourse, compared to P1 and P2, except for *endophoric markers*, which were used more frequently in P2.

4.2.2.3 Patterns of stance markers in identity construction in Nancy's papers

As exhibited in Table 15, stance markers were rather evenly distributed across Nancy's three papers. In total 25.9% of the metadisourse items were stance markers, accounting for 10.4 times per 1,000 words. This pattern showed that in general, Nancy was a writer who would take up positions and express judgements in relation to the propositions, results and/or research phenomena in her academic papers.

The most frequently occurring stance marker category, in all three papers, was *hedges*. Though the frequency (after normalization) was pretty close, hedges appeared the most frequently in P2, followed by P3 and P1. The pattern showed that in P2, report written for state government, Nancy (and the research team) displayed their identities as having more deference and reducing the tone of certainty when writing for their direct sponsors. *Boosters* were used less than *hedges* in Nancy's papers, and this category was found the most frequently in P3, followed by P2 and P1. *Attitude markers*, on the contrary, were used the most frequently in P1, and least frequently in P3. Note that Nancy's three papers were written across different times of her PhD studies; this trend showcased that diachronically, Nancy's employment of *boosters* increased, while the use of *attitude markers* decreased. *Boosters* used more in P3 probably demonstrated that as a PhD candidate who was about to present the concluding project to the examiners, Nancy was constructed as a writer who assumed more confidence and self-assuredness in presenting her research and to establish herself as a qualified degree winner. The decreasing pattern of *attitude markers*, however, might show that Nancy gradually was constructed as an author who reduced the subjective tone and presented the information in a more objective way.

Last, *self-mentions* did not appear extensively in Nancy's papers, only 0.3 times in P1 and 0.8 times in P3 after normalization. It is notable that in P2, no instances of *self-mentions* were found. This demonstrated that in P2, the government report, individual voice probably was not critical. Nancy's identity was only a presenter of information collected and analyzed. Her personal agency as a writer, was downplayed in this genre.

4.2.2.4 Patterns of engagement markers in identity construction in Nancy's papers

As demonstrated in Table 16, in general, engagement markers were used negligibly in Nancy's papers. Only two categories of engagement markers appeared in all the three papers, including *directives* and *appeals to shared knowledge*, with the former one being the most frequently used in the three papers. Though *directives* were all used at a rather low frequency, they were used more frequently in P3 than in P2 and P1. As to *appeals to shared knowledge*, slightly more instances were found in P3. In regard to *reader pronouns*, they were only used in

P2, the report written for the government. *Questions*, on the other hand, only appeared in P3, the dissertation. Last, *personal asides* were only used once in P3.

The extremely low frequency of engagement markers showed Nancy's identity as a writer who did not engage her readers extensively in her papers of different genres. By using *directives* in all the three papers, Nancy was constructed as an agent of her research who then was qualified to instruct readers in her preferred ways. *Appeals to shared knowledge,* appearing in all the papers, showed that Nancy constructed a disciplinary insider identity and aligned herself with her readers by assuming that the readers shared the knowledge in her papers.

4.2.3 Factors shaping Nancy's identity construction in the three papers

There were various factors that contributed to Nancy's identity construction in the three papers. Those related to *autobiographical self* included her personal writing habits, previous training regarding English academic writing, and her personal experience. *Possibilities for selfhood* influenced Nancy's identity construction in terms of advisors' feedback, reviewers' comments, and writing conventions in the field. She also exerted agency in regard to metadiscoursal choices to construct the intended identities. These factors are elaborated more closely in this section.

4.2.3.1 *Autobiographical self* and *discoursal self* and *self as author*

4.2.3.1.1 Writing habits

An individual's identity construction can be highly related to their idiosyncratic writing style (Ivanic, 1998). In Nancy's case, her personal writing habits influenced her identity construction, which could be reflected by some of her metadiscoursal choices. In terms of *self-mentions*, Nancy used "we" in her dissertation to refer to herself as an author, which she stated was due to her writing habits. According to her, this was a “bad writing habit”, and she should have used “I” in her dissertation to claim authorship. As our conversation went,

> Nancy: (Using “we”) is my writing habit, a bad writing habit. Though my committee members didn’t point this out, I knew from somebody else’s dissertation defense and the members said “I” should be used. This is a bad writing habit.
>
> Ren: What makes you feel it’s bad?
>
> Nancy: Because from the education I had, academic writing is not somewhere you can voice your subjective opinions. Maybe you can use more objective expressions, like passive voice or “it is…”. Using “we” probably isn’t good. This is something I can continue working on in my English academic writing.
>
> Ren: So, you think you should have used “I” in your dissertation instead of “we”.
>
> Nancy: I totally agree. (I3)

However, the habit of using “we” seems to not apply in the P2, the research report, which had a specific group of researchers carrying out the study. In P2, Nancy did not use “we”; instead, she used either “the research team” or passive voice to downplay the roles of self-reference pronouns. As she said in our interview, not using *self-mentions* in the report was also because of her writing habits:

> Ren: Here you used passive voice, such as “a questionnaire was developed”, etc. When you described something you did, you used passive voice. What kind of considerations did you have when writing that way?
>
> Nancy: This is my writing habit. Actually, there’re not many rules about that. In report, you can say “our research team did this” or passive voice.

Ren: You also used plenty of "the research team did this and that"; you didn't use self-referenced pronouns such as "we" or "I" did this and that. Does that have to do with the report genre?

Nancy: I don't think so. I think it's purely my writing habit. (I1)

From Nancy's responses, it is clear how her personal habits played a critical role in her identities she probably was not conscious of constructing in different genres. This inconsistency seems interesting and it will be further discussed in the discussion chapter.

4.2.3.1.2 Previous training on writing in English

Previous training on English academic writing, particularly which Nancy had in China, still had profound effect in her current academic writing and that contributed to her construction of identity. As the results reported in the previous section show, Nancy only used *self-mentions* "we" twice in her P1. According to her, she tended to avoid using self-referenced pronouns and to use passive voice instead in order to reduce the subjectivity that might be caused by using self-mentions. Nancy stated that this tendency to avoid bringing herself to the fore could be traced back to the training she received in her undergraduate program. As has been provided in Nancy's profile, Nancy completed her bachelor's degree in a bilingual program in China. The program offered a course related to business writing, from which she learned that "we" should not be extensively used in academic writing because it may sabotage the objectivity. She said:

> I have been using passive voice. I know this writing style may not be good, but I've been using it. This can make the papers less subjective. (I1)

When consulted about the reasons for this preference of avoiding using *self-mentions*, Nancy answered:

> Maybe this is related to the earliest training I received… My writing style is that I barely use "we". I use passive voice more… One reason may be that in my bilingual program, I had a course which was something about business writing. My teacher said how to avoid using "we" all the time in the paper. I just feel being so subjective in academic writing may make readers uncomfortable, so I try to avoid it (*self-mention*)… From then on, I have been pushing myself to use passive voice and this writing preference has been retained till now. And in my doctoral program, it's never been an issue from the feedback my advisors gave me. (I1)

Nancy occasionally used *attitude markers* to express her affective stance in her papers. As has been mentioned in Nancy's profile in the methodology chapter, Nancy had taken IELTS and GRE courses when she was in China. She remarked that using adjectives to signal the attitudinal stance she assumed toward results or propositions in her papers was caused by her prior experience related to studying IELTS. This influence can be reflected in the following interview extract:

> Nancy: I often use (adjectives). I can't say that they are very common, but I use them very often, because I've been using those words throughout my English learning and writing experience related to graph/figure description. I always write this way.
>
> Ren: Okay, graph/figure description, like what you learned in your IELTS or TOEFL courses.
>
> Nancy: Yeah, it's the repercussion of me studying IELTS. (I3)

Nancy's experience of studying for the tests (IELTS and GRE) in China to apply for an overseas PhD program influenced her English academic writing not only in terms of *attitude markers*. In her use of *hedges*, she tended to strategically use modal verbs for the purpose of avoiding using third person singular or past tense, grammatical structures she considered relatively easier to make mistakes. In the interview (I3) for her dissertation, she said:

> Well… I'm concerned that I'd make mistakes if I used third person singular and past tense. So, I often used "may" in my writing… Using "may" can avoid mistakes. And this is the trick I learned in my GRE writing courses. My colleague back then told me that this is a trick. It can help avoid a lot of grammatical mistakes. If you make mistakes on third person singular, you'll definitely lose points in your writing. But using "may" would avoid that. (I3)

That Nancy tried to avoid making mistakes through the use of the *hedges* such as "may" and "might" was also reflected in our interview on her P1. I asked her the reasons for using these *hedge* items (may and might) rather densely in the introduction section, and she answered,

> It's easy to make mistakes on third person singular and tense. But in a lot of situations, you don't really need to make things clear through writing a sentence with verb tense… I use a lot of "can" and "may", because this can avoid mistakes related to third person singular and tense. (I1)

4.2.3.1.3 Personal experience

In Nancy's dissertation, it was found that she used *appeals to shared knowledge* as indicated by such expression as "common sense" in the following sentence

> This conclusion is consistent with two previous studies (Wang et al., 2017; Zhao et al., 2015) and the <u>common sense</u> that drivers need more time to adapt changes of high-to-low or low-to-high lighting patterns… (P3)

The statement made, according to Nancy, was influenced by her personal experience, as she explained:

> Normally, if I use common sense, that's because I try to use my knowledge, my domain knowledge and my view of the world to understand my model and the results. I didn't mean my model told me it needs more time; it's my life experience that told me so. So I'd use that word. (I3)

4.2.3.2 *Possibilities for selfhood* and *discoursal self* and *self as author*

4.2.3.2.1 Advisor's feedback

Nancy's advisor's influence on her identity construction noticeably revolved around the use of *transitions*, which contributed to the presentation of a scholarly and competent academic writer persona. As has been reported in the previous section, Nancy used *transitions* roughly at the same frequency in her P1 and P2 (11.9% vs. 11.5%), whereas in P3, her dissertation, more *transitions* were found (17.2%). It was also mentioned that Nancy's dissertation entailed three separate studies, the first two being her completed projects that had been published, whereas the first study (Chapter 2) was a revised version of P1. The third study which she started from scratch (Chapter 4) during the dissertation writing stage was found to have more *transition* markers. I brought up this difference in our interview and she explained that this was mainly because her dissertation advisor had given her guidance particularly in this regard, who strongly advocated the importance of the internal flow and logic of academic writing.

> Ren: A difference I found between the three chapters of your dissertation is that the third study obviously used more *transitions* than the other two studies. So that's about the flow you just said.

Nancy: The flow is very smooth, right? Just smooth. You'll feel that there's no breakup point… You won't find that after I'm done with talking about AB and C, then I come back again to talk about B. The *transitions* are to make the connections smooth. (I3)

As has been specified in the methodology chapter, Nancy had two advisors, one being the one in the research center she worked in, and the other being the professor from the doctoral program she belonged to. The projects she undertook when working on her PhD were mainly collaborations with the advisor and co-workers from the research center. Her advisor from the engineering program basically only guided her writing when she was working on her dissertation. Nancy voiced that using *transitions* more was due to the dissertation advisor's specific writing guidance.

4.2.3.2.2 Reviewers' comments

Reviewers' comments played a crucial role in helping shape Nancy's identity in relation to metadiscourse usage. In Nancy's P1, one of the anonymous reviewers specifically pointed out several important previous studies Nancy failed to review them in the introduction section, as shown in the following excerpt:

Introduction

The authors fail to mentioned *[sic]* three major prior studies that evaluated the impact of street lighting photometric measures on nighttime crashes. The authors need to critically assess how the results of the following studies agree or disagree with results of their study. The papers are mentioned as follows:

1. Gibbons, R., Guo, F., Medina, A., Terry, T., Du, J., Lutkevich, P., & Li, Q. (2014). Design criteria for adaptive roadway lighting (No. FHWA-HRT-14-051).

2. Bhagavathula, R., Gibbons, R. B., & Edwards, C. J. (2015). Relationship between roadway illuminance level and nighttime rural intersection safety. Transportation Research Record: Journal of the Transportation Research Board, (2485), 8-15.

3. Edwards, C. J. (2015). Lighting levels for isolated intersections leading to safety improvements.

(Extracted from journal's blind reviewers' feedback on P1).

As shown, the reviewer offered specific feedback in regard to how to make improvements in the use of *evidentials*. Through integrating the reviewer's critical comments in her revised paper, Nancy could draw upon *evidentials* from relevant studies. Another reviewer also mentioned the issues existing with the use of citations in Nancy's P1, as shown in the following excerpts:

Lines 10-12: Provide a reference to substantiate the statement that reduced visibility, drowsiness and impairment are what contribute to nighttime crashes.

Lines 48-49: Provide a reference to support the statement that the NB model can overcome overdispersion commonly found in crash data.

(Extracted from journal's blind reviewers' feedback on P1).

The specific revisions requested by the reviewers with regard to *evidentials* prompted Nancy to present relevant information to support the arguments, which would strengthen the credibility of the propositions and would contribute to her identity construction as a disciplinary insider and a knowledgeable academic writer.

As said in the methodology chapter, Nancy's advisor from the research center hired a native speaker of English to proofread their papers. The proofreader mainly focused on the format, structure, and spelling and other technical issues. In the proofreader's feedback, however, she did point out problems in Nancy's report regarding the presentation of Figures and

Tables. For example, as shown in the following Figure, through suggesting the possible absence of some information, proofreader's comments helped Nancy notice this problem and promote the *endophoric markers* used in the report. Making improvement in this regard could facilitated Nancy's identity construction as a competent and professional report writer.

Figure 7. Comments on Nancy's P2 from the proofreader

4.2.3.2.3 Writing conventions

Writing conventions were essential for writers to follow in their academic papers, which could provide the repertoire for writers to choose from and construct their identities. Writing conventions from the disciplinary and academic context influenced Nancy's identity construction related to metadiscourse use in several ways. The first was *evidentials* in Nancy's papers. As presented in the methodology chapter, the three papers Nancy sent me were written for different purposes. In her P2, the government report, the sources were referred to as numbers in parentheses, while in P1, the paper submitted for a journal, citations included authors' names and year of the work that was published. When asked about this difference in regard to *evidentials*, Nancy stated that this was required by the journal:

> Ren: I noticed some differences in your use of citations in different papers, such as listing them as numbers in brackets vs. offering names and year in brackets.

Nancy: That was journal's requirements…I don't know about your discipline. In mine, journals' requirements are pretty detailed… (I1)

The institution Nancy was from also had specific rules on how citations should be attributed. Nancy's use of *evidentials* in her dissertation (P3), therefore, was performed according to the authoritative rules, as she said:

Because this part (citations) actually is university's requirements. It has unalterable rules. No one can customize it… I think it should be the requirements from the dissertation website… (I3)

Not only did the writing requirements from authoritative multiple institutions shape Nancy's identity construction, the conventions of writing for different purposes also influenced Nancy's identity construction in regard to the use of *evidentials* as well. As presented in the previous section, citations in Nancy's three papers showcased great differences (6.1 times per 1,000 words in P1, 2.4 times per 1,000 words in P2, and 10.3 times per 1,000 words in P3), with *evidentials* appearing less frequently in report than in the other two genres. Nancy specified the differences in regard to citation use in the different types of her writing:

When I write literature review, it's like, when you write papers, you have to clarify a problem, so you review a lot of studies and analyze the issue layer by layer. But for literature review in report, especially like this one that's implementable, we just review those studies that are very similar, literally those that used the same methods and data as the report did. (I2)

Writing for different purposes influenced Nancy's other metadiscoursal choices that contributed to her identity construction as well. As has been reported in the previous section, *endophoric markers* were used significantly more in P2, the government report, than in P1 and

P3. In P2, the metadiscoursal marker appeared 5.5 times per 1,000 words, whereas it appeared 1.9 times in P1 and 3.1 times in P3. This was due to the different writing requirements of the three writing tasks. In Nancy's P2, the purpose was to present to the project manager who funded the research what the research team had done. The final report, thus, synthesized the efforts the research team had put in, as Nancy pointed out the differences between a report and a journal paper:

> (For report) your audience is, for example, your project manager. He knows what you've done every time he listens to your presentations. On the other hand, your audience is people who give you fundings. They have to know that the $200,000 they give you is not wasted. So you do something, and you present in the report what you do and how it can be used. That's fine. But papers are different. You have to tell people that what you're doing is unique and new. You make contributions. (I1)

From Nancy's analysis of the purposes of report, it was clear that reports focused more on the description and presentation of the experiment conducted and information gathered by the research team. A distinctive feature that could be found in Nancy's report was that a considerable number of Tables and Figures were used (64 Tables and 13 Figures in the report). To refer to those multimodal information, Nancy used *endophoric markers* extensively to explain the information, as she said:

> I think… report is not that… because report mainly highlighted what I did… For engineering practices, Tables, Figures, and Graphs are more visually intuitive. In terms of writing, probably you cannot find changes in my wording; you can only see how I organize things. (I1)

The writing conventions also shaped Nancy's identity construction in terms of the use of *endophoric markers*. She stated that one of her preferences for describing data was to present them in a Table, and she imitated this from an established scholar in her field, as she said:

> Like in my papers, you can find I'd insert a Table describing all the data I use. This actually was me imitating a big name's style in our field. The papers he published decades ago, 30 papers of his were in this structure. His advisees, who are also very successful researchers in our field, also write this way. So, I just "go with the trending" (original phrase from Nancy). It's also easy for people to understand. (I1)

The report written for government also was heavily influenced by the rules writers were expected to follow to perform this genre. For instance, in the report (P2), there was a part named "executive summary", which functioned as a *frame marker* summarizing the methods and findings of the report. Nancy explained she wrote this way to strictly follow the template:

> Yeah, about the structure, actually it's from the template. We do consulting for XX (the government department), and their report has a predefined structure. We have to follow it, because they are our client. (I2)

Similarly, *endophoric markers* were used in Nancy's report extensively to summarize the main content at the end of a section. This, according to her, was also a convention prescribed in the report template:

> Ren: I noticed that at the end of each section or chapter, you'll have a summary, covering the methods you used, what conclusions you got. Like you'll mark some highlight. Is this a convention in writing report?
>
> Nancy: These are all from the template. I personally think report is too wordy. (I2)

Endophoric markers were used in Nancy's report in the beginning of a section or chapter to specify what would be covered in that part. Nancy addressed that this writing convention was also from the template, which could direct readers' attention and organize the text.

> Ren: Like in here you have sentences such as "these studies are summarized in the following sections based on their objectives…", like you're telling readers what's I'm gonna say next. Is this part of the template?
>
> Nancy: Yes, it's template… The template can be considered an outline. It has rules you need to follow in different sections. It's like a dissertation. In different sections you need a summary to lead your readers. (I2)

4.2.3.3 Agency and *discoursal self* and *self as author*

4.2.3.3.1 Clarification of information

Nancy's agency played an essential role in the construction of her identities in the papers she wrote. In her dissertation, she employed *code gloss* extensively to offer explanations in parentheses in order to make the information more comprehensible and self-evident. When asked about the reasons behind this noticeable metadiscoursal use, she answered,

> I don't know if this is a good or bad habit, but I think there's some logical reasoning I didn't make very clear (in the sentences). I feel it (the content in the parentheses) is a "lost part". It's not like it can follow the sequence of "ABC", and can be connected with what comes before and after the explanation. So, I put them in the parentheses so I can make it clearer. Yeah, I feel I can't make the logic smooth if it's written in the sentence, so this is the only thing I can do. (I3)

4.2.3.3.2 Nature of content

When offering explanations and interpretations, it was found that Nancy often used *hedges* in her papers. Besides using *hedges* for avoiding grammatical mistakes, as reported in previous section, she was aware that the use of this was also dependent on the nature of content. She had actively chosen from the available words to fulfill this rhetorical effect. As she stated the reasons for her consciously selecting *hedges* in P1,

> Like this research I did, it describes the results of a model. Many reasons are unknown to us. So I can only use words like "can, may, possible, likely"… You have to use your "domain knowledge" to explain the reasons for that. These reasons are possible reasons, not the only reason. So I will use words like "may, can". (I1)

Nancy used *attitude markers* in her papers to express attitudes toward certain propositions or research results. Regarding this, she consciously selected those adjectives or the structures (it is + adjectives) to promote grammatical and syntactical diversity in her academic writing. When I asked her about the frequently found structure (it is + adjectives), she said:

> Ren: I found you used a lot of words such as "it is important to note that…"
>
> Nancy: That's a sentence structure I often use… Yes, you cannot use the same structure "it is…" or "I blahblah" throughout the whole paper. I use this just to make my paper less monotonous. (I2)

For the same purposes, she used *boosters* together with *attitude markers* in some situations in P2, as reflected in our conversation:

> Ren: I noticed that you used words like "it is strikingly clear…"
>
> Nancy: That's just an adjustment to my sentence patterns.

Ren: Have you thought about what kinds of effect it will attain? Or how will your readers respond?

Nancy: Sorry, I've never thought about that. Literally I just take it as an adjustment in my writing. (I2)

When reporting results, Nancy also inserted her judgements of some results by using the attitude marker "it is interesting to find that…" multiple times. When asked her opinions about expressing her own evaluations through the adjectives, she insisted that adding diversity to the sentence structure was the only reason for her use of this metadiscoursal item:

> It's really just an adjustment. I don't really think it's interesting, to be honest. (I2)

Nancy then explained the reasons for using the attitude markers for its own sake. She actually signaled her affective stance (even though she stated she did not really think in those ways) due to the robotic style of writing a report, which was supposed to strictly follow the template issued by their sponsor, and she wanted to make the report less monotonous. As Nancy said,

> Those sentence structures are my only personal expressions in the report. I just want to make it look a bit different, because I know in report you'll always write in the same way, like "after we did this, after we did this, after we did this". I'm really resistant of that… So to make it less boring, I have "it is clear…" etc. so that there'll be more diverse expressions. More diversity. (I2)

It is interesting to find out the conflict and inconsistency between Nancy's metadiscourse choice and her lack of consciousness of how those linguistic choices would influence her identity construction. This will further be discussed in the last chapter.

4.3 Adam

In this section, I report findings related to Adam's identity construction through metadiscourse in his three papers, starting with the overall distribution of metadiscourse markers, followed by the patterns of metadiscourse use in identity construction in his papers, and the factors influencing his identity construction in the three papers.

4.3.1 Metadiscourse usage in identity construction in Adam's papers

In this section, I present how Adam's identities were constructed through the use of metadiscourse, first through the overall distribution of metadiscourse in his papers. Then I report the distribution of each category of metadiscourse starting with interactive metadiscourse, then stance and engagement markers under interactional metadiscourse.

4.3.1.1 Overall distribution of metadiscourse markers in Adam's papers

As demonstrated in Table 17, altogether there were 755 cases of metadiscourse markers in Adam's three papers, which was 48.4 times per 1,000 words. Interactive metadiscourse markers were overwhelmingly used in Adam's three papers, with 613 cases present, covering 81.2% of the total metadiscourse markers. After normalization, interactive metadiscourse appeared 39.3 times per 1,000 words. Stance markers were found to be used 120 times in Adam's data, which covered 15.9% of the total metadiscourse markers, constituting 7.7 times per 1,000 words. Lastly, engagement markers only were used 22 times, covering 2.9% of all the metadiscourse markers, and amounting to 1.4 times per 1,000 words.

In Adam's P1, the book review, 84 metadiscourse markers were found, which was 50 time per 1,000 words after normalization. More specifically, 62 of them were interactive metadiscourse, covering 73.8% of the total metadiscourse markers in this paper. Next, stance markers were used 19 times, covering 22.6% of the total metadiscourse markers. Engagement markers appeared 3 times in P1, covering 3.6% of the total metadiscourse markers.

The P2 of Adam's, the paper about university websites, was a course paper for a final project from a course Adam took in the first semester of his PhD. In this paper, 321 cases of metadiscourse were found. A majority of them (260 items) were interactive metadiscourse, covering 81% of total metadiscourse markers, which equaled to 35.8 cases per 1,000 words. Stance markers appeared 52 times, constituting 16.2% of the total metadiscourse markers, which was 7.2 times after normalization. Engagement markers, however, appeared 9 times in P2, covering 2.8% of the total metadiscourse markers, equaling 1.2 times per 1,000 words.

Table 17. Overall distribution of metadiscourse markers in Adam's papers.

	P1		P2		P3		Total	
Category	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words
Interactive metadiscourse	62 (73.8%)	36.9	260 (81%)	35.8	291 (83.1%)	43.7	613 (81.2%)	39.3
Interactional metadiscourse (stance)	19 (22.6%)	11.3	52 (16.2%)	7.2	49 (14%)	7.3	120 (15.9%)	7.7
Interactional metadiscourse (engagement)	3 (3.6%)	1.8	9 (2.8%)	1.2	10 (2.9%)	1.5	22 (2.9%)	1.4
Overall total	84 (100%)	50	321 (100%)	44.2	350 (100%)	52.5	755 (100%)	48.4

Last, Adam's P3, another final project submitted for a course in the first semester of his PhD, encompassed 350 cases of metadiscourse markers, equaling to 52.5 times per 1,000 words. Among them, 291 cases of interactive metadiscourse were found, covering 83.1% of all the metadiscourse markers, which was 43.7 times per 1,000 words. Stance markers were used 49 times, covering 14% all the metadiscourse markers, accounting for 7.3 times after normalization. Engagement markers appeared 10 times, constituting 2.9% of the metadiscoursal resources, which was 1.5 times after normalization.

4.3.1.2 Interactive metadiscourse and identity construction in Adam's papers

Table 18 displays the distribution of interactive metadiscourse markers in Adam's papers. As shown, this category was used predominantly across the three papers, covering 73.8%, 81% and 83.1% of the total metadiscourse markers in P1, P2 and P3 respectively. The dominance of interactive metadiscourse underlay Adam's identity as a writer who employed explicit linguistic signals to keep texts organized and keep readers informed in regard to where they are and where they are going (Hyland, 2018).

Transitions were the category of interactive metadiscourse that was used rather frequently across Adam's three papers, covering 40.5%, 22.7% and 23.1% in P1, P2 and P3 respectively. After normalization, they appeared 20.2 times, 10.1 times and 12.2 times in the three papers. As explained earlier, *transitions* are essential in academic written discourse so that the relations among sentences could be specified, and could facilitate readers' processing of the texts in a logical manner. Adam used various kinds of transitional markers to link his text together, as shown in the following examples:

> <u>On one hand</u>, Bilibili is a home for fans of Japanese animations and manga, <u>so</u> their online literacy practices would be influenced by Japanese culture. Weibo, <u>on the other hand</u>, was used for idol fandoms influenced by K-wave, <u>so</u> the online practices would be affected by Korean language and culture. (P1)
>
> <u>However</u>, the age of authority does not die, <u>and</u> the distinction is shown in a covert and naturalized way via discourse strategies (Vincent, 2018). (P3)

> Likewise, *About us* of top-tier universities prefers more informal information and students' evaluation, even though the websites maintain their alignment with governments… (Xue et al., 2021). In addition, top-tier and second-tier universities' websites … were compared by Xie and Teo (2020). (P2)

As the instances illustrated, Adam's use of *transitions* was fairly densely distributed in his papers. In so doing, Adam specified the relations between propositions based on his understanding and his preferred ways of organizing information. The high frequency also showed that he did not totally rely on readers to interpret the underlying relations or structures of his texts. The extensive use of *transitions* could possibly help promote readers' recovery of the communicative intentions (Hyland, 2018) more effectively. Thus, Adam's use of *transition* markers constructed himself as a competent academic writer, who addressed readers' needs and ensured that readers could draw his intended connections.

Evidentials were another category that were used quite frequently in Adam's papers, covering 7.1%, 25.6% and 27.1% of metadiscourse markers in the three papers, which equaled to 3.6 times, 11.3 times and 14.3 times per 1,000 words respectively. As a distinctive feature in academic discourse, *evidentials* assumed a major position in Adam's papers. They tended to appear frequently in cluster, as the following instances demonstrate:

> As social media permeates every aspect of life and has become a necessary component of our socialization, people already entered post-digital era (Lyons et al., 2021). Different from the Web 1.0…, social media, the representation of Web 2.0 is characterized by interaction, and multi-authorship and multimodality (Tagg, 2015). (P1)

> Previous studies pertinent to the marketization of Chinese higher education revealed the notion of interdiscursivity, implying that different kinds of discourse were mixed (e.g., Han, 2014; Teo & Ren, 2019; Feng, 2019; Xiong, 2012; Xiong & Li, 2020). (P2)
>
> It can invoke diverse identities and ideologies, such as gender identities (Cairns, Johnston, & Baumann, 2010) and ethnic identities in multiethnic societies (Reddy & van Dam, 2020). According to Bourdieu (1984), though the overt social hierarchy based on the social position or wealth becomes less acceptable in societies.... (P3)

Note that Adam's three papers were submitted to two courses in his first year in the PhD program to fulfill course requirements. Therefore, using citations in his papers, on the one hand, could demonstrate the scope of his field knowledge to the reader (aka. his instructor) and potentially showcase the sufficient work he had done, which could help him gain a good impression. On the other hand, the sources Adam cited could construct his writer's identity that was interested in a specific field, which may be significant for Adam to do so. As mentioned in the methodology chapter, the reader was a researcher sharing similar research interests with Adam and who Adam wanted to work with in the days to come. To construct this identity was important at the beginning of his PhD studies, as this would influence the directions Adam planned to pursue and who the professor/s would be that he aimed to work with.

Code gloss was used rather densely in Adam's papers, constituting 16.7% of the total metadiscourse in P1, 19.3% in P2, and 20.9% in P3, which were 8.3 times, 8.5 times and 10.9 times in the three papers after normalization. *Code gloss* markers were crucial to provide further explanations regarding concepts or entities, and they could appear in different forms, such as being a part of a sentence or within parentheses, as displayed in the following examples:

> In China, platforms <u>such as</u> Weibo and Bilibili, share the similar function as Twitter and YouTube respectively (Wang, 2016). (P1)
>
> Japanese expression, <u>say, 胖次 (pangci, lit fat guy)</u> referring to a Japanese word, underwear, due to the similar pronunciation. (P1)
>
> …<u>such as</u> 2+2 programs <u>(students spend two years in Chinese universities and two years in foreign universities)</u> and 3+1 programs <u>(three years in China and one year in foreign countries)</u>. (P2)
>
> Tongxiemen <u>literally means</u> children' shoes but it can refer to classmates <u>(tongxuemen)</u> online as a general addressee. (P3)

As demonstrated in the methodology chapter, Adam's three papers directly or indirectly involved data from the Chinese context. Hence in the papers, there were many situations he needed to explain what the original concepts or words meant in English in order for the readers to correctly understand the intended meanings of his analysis. As the examples revealed, Adam reformulated in an approachable way in English some expressions or cultural phenomena that might be new to the readers. This constructed Adam's real-world identity as a second language writer whose native language was Chinese. On the other hand, use of *code gloss* in the papers demonstrated Adam's identity as a considerate writer who was aware of readers' needs to understand the content properly, and who tried to write in a context-specific manner to cater to readers' needs.

Table 18. Interactive metadiscourse in Adam's papers

	P1		P2		P3		Total	
Category	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words
Transitions	34 (40.5%)	20.2	73 (22.7%)	10.1	81 (23.1%)	12.2	188 (24.9%)	12
Frame markers	8 (9.5%)	4.8	10 (3.1%)	1.4	6 (1.7%)	0.9	24 (3.2%)	1.5
Endophoric markers	0 (0%)	0	33 (10.3%)	4.5	36 (10.3%)	5.4	69 (9.2%)	4.4
Evidentials	6 (7.1%)	3.6	82 (25.6%)	11.3	95 (27.1%)	14.3	183 (24.2%)	11.8
Code gloss	14 (16.7%)	8.3	62 (19.3%)	8.5	73 (20.9%)	10.9	149 (19.7%)	9.6
Interactive (Total)	62 (73.8%)	36.9	260 (81%)	35.8	291 (83.1%)	43.7	613 (81.2%)	39.3

Next, *frame markers* were used comparatively less than the aforementioned three sub-categories of interactive metadiscourse markers. They constituted 9.5% of the total metadiscourse markers in P1, 3.1% in P2, and 1.7 in P3, and was 4.8 times, 1.4 times and 0.9 times respectively in the three papers after normalization. As mentioned earlier, *frame markers* can be divided into four categories: announcing goals, sequencing, labelling stages, and shifting topics (Hyland & Jiang, 2022). However, only two types were found in Adam's papers: sequencing and labelling stages. Adam tended to use sequencers such as "firstly" and "secondly" to list the components of an entity, as in the following two examples:

> In this section, the author provided in-depth analysis of the findings. First, he argued that Chinese people can use diverse linguistic and non-linguistic semiotics to construct meanings. Second, the use of diverse semiotic resources was a process of translocalization and recontextualization. (P1)

> There are two research gaps. Firstly, the research of digital food review was mainly studied in English context and on ‘western’ platforms… Secondly, as mentioned before, food reviews on Instagram were characterized as more pictures and fewer texts (Feldman, 2021). (P3)

As the examples demonstrate, Adam clarified the information (the findings and the research gaps) layer by layer, which could, first, make the sophisticated information more organized, and second, helped orient and guide readers through the information (Hyland, 2018).

Frame markers were found to function to label stages in Adam’s paper. They were often found toward the end of a section to summarize what had been covered in this section. As the examples showed:

> In conclusion, the dissertation is well structured to facilitate social media researchers to read through. (P1)
>
> In a word, various semiotic landscapes on the websites have manifested the ideologies about marketization, cultural contexts and educational policies in different countries. (P2)

Using “in conclusion” as well as “in a word” to perform the discourse act at a particular point, Adam indicated the stages his argumentation had arrived at. Using *frame markers* to attain the functions of sequencing and labelling stages, Adam was constructed as an academic writer who wrote in an organized and reader-friendly way, and who used signposts to help readers to understand the materials.

Lastly, *endophoric markers* appeared only in P2 and P3, with no cases found in P1. In these two papers, they both covered 10.3% of the total metadiscourse resources. After normalization, in P2, they appeared 4.5 times per 1,000 words, and 5.4 times in P3. *Endophoric markers* pointed to information in other parts of a paper. As mentioned, Adam’s papers examined platforms in Chinese cyber communities; hence multimodality, such as images could be found in his papers to present the information under investigation. As the first three examples below demonstrate, Adam used *endophoric markers* to locate the images and/or data and made corresponding explanations:

> In Figure 6, a student and a white teacher are working together on the student's assignment. (P2)
>
> An emoji of a fairy in excerpt 4 is to emphasize the argument of the reviewer. (P3)
>
> Figures 1 and 2 are the two templates of showing how international universities guide viewers to read each characteristic. (P2)
>
> Secondly, as mentioned before, food reviews on Instagram were characterized as more pictures and fewer texts (Feldman, 2021). (P3)

Other than referring to images, Figures or Tables, Adam also employed *endophoric markers* to refer to information presented elsewhere in the papers, as shown in the fourth example above. *Endophoric markers* constructed Adam's identity as a writer who demonstrated his field knowledge in presenting the analysis, following the ways commonly accepted by the academic conventions from a broader circle.

4.3.1.3 Stance markers and identity construction in Adam's papers

Table 19 displayed the distribution of Adam's use of stance markers in his three papers. *Hedges* were the category that was very frequently used, which covered 7.1% of all the metadiscourse markers in P1, 9.7% in P2, and 8.6% in P3. After normalization, hedges appeared 3.6 times in P1, 4.3 times in P2, and 4.5 times in P3 in every 1,000 words. As the following examples demonstrate, the use of modals and the phrasal forms represented Adam's explicit intrusion into the text and conveyed his personal assessment on the propositions, which indicated the cautious writer identities he constructed were based on his plausible reasoning instead of certain knowledge (Hyland, 2012b):

…so the online practices <u>would be</u> affected by Korean language and culture. (P1)

While C9 universities all state that they are the first universities in this country, international universities <u>seem not to be</u> compatible with others… (P2)

The social semiotic approach to multimodality <u>tends to</u> analyze diverse affordances, including hashtag, images, and emojis. The social semiotic regarded that each sign/mode <u>could have</u> three functions… (P3)

The national or regional heritage identities <u>tend to be</u> constructed to emphasize their expert knowledge in their national food (Vásquez & Chik, 2015). (P3)

Table 19. Interactional metadiscourse in Adam's papers – stance

	P1		P2		P3		Total	
Category	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words
Hedges	6 (7.1%)	3.6	31 (9.7%)	4.3	30 (8.6%)	4.5	67 (8.9%)	4.3
Boosters	8 (9.5%)	4.8	13 (4%)	1.8	12 (3.4%)	1.8	33 (4.4%)	2.1
Self-mentions	0 (0%)	0	2 (0.6%)	0.3	5 (1.4%)	0.7	7 (0.9%)	0.5
Attitude markers	5 (6%)	2.9	6 (1.9%)	0.8	2 (0.6%)	0.3	13 (1.7%)	0.8
Interactional - stance (Total)	19 (22.6%)	11.3	52 (16.2%)	7.2	49 (14%)	7.3	120 (15.9%)	7.7

Boosters were used in Adam's three papers, which constituted 9.5% of all the metadiscourse markers in P1, 4% in P2, and 3.4% in P3. The frequencies were 4.8 times, 1.8

times and 1.8 times per 1,000 words in the three papers respectively. From the following examples, it is clear that Adam tended to use adverbs and emphatics to amplify the tone and to highlight some information, as in:

> The author stated the importance of context and nature of multimodality of social media for data analysis, <u>especially</u> for the bullet curtain on Bilibili… (P1)
>
> The paper <u>showed</u> how marketization was reproduced on the websites of Chinese universities and unpacked the homogeneity of Chinese universities. (P2)
>
> The successful further studies or career choices of graduates also <u>proved</u> the credibility and qualification of international universities. (P2)
>
> The gender identity is also constructed multimodally, <u>especially</u> by selfies. (P3)

By using *boosters*, Adam strengthened the credibility of his study, and the use of boosting expressions constructed himself as a confident academic writer "with a strong opinion to uphold, and the right to express it" (Clark & Ivanic, 1997, p. 150).

Attitude markers, on the other hand, were used less than *hedges* and *boosters*. The frequencies were 6% in P1, 1.9% in P2 and only 0.6% in P3. After normalization, the frequencies were 2.9 times, 0.8 times and 0.3 times every 1,000 words. Upon further looking at the data, it was found that Adam preferred to use "it is + adjectives" to express his affective stance toward the propositional entities in his papers, as the following examples demonstrate:

> <u>It would be better that</u> authors could explore socio-cultural factors of the two platforms. (P1)
>
> <u>It is not surprising that</u> the history of international universities is shorter than the C9 league. (P2)

In addition, it is worth noting that the nominal,米其林小馆子, literally translated as

Michilin little restuant *[sic],* is used by the reviewer. (P3)

Adding Adam's own personal opinions displayed his affective positions toward the propositional content. This way, he was constructed as an author who demonstrated his "involvement with colleagues by positioning oneself in terms of disciplinary values to express a community-ratified stance" (Hyland, 2012b, p. 147).

Self-mentions only occurred in P2 and P3, the two research articles for final course projects. In P2, the appearance of *self-mentions* accounted for 0.6% of the total metadiscourse markers, which was 0.3 times per 1,000 words, whereas in P3, self-mentions covered 1.4% of all the metadiscourse items, 0.7 times after normalization. It needs to be pointed out that Adam exclusively used "the author" in the two papers to refer to himself. For example:

> The author chose the mode of popularity and collected the top 100 posts. (P3)
>
> As for the problem whether they are willing to be involved in this study, the author referred to Page's (2018) argument, since the social media platforms are semi-public sites… (P3)
>
> The author argued that the locations such as Shenzhen and Shanghai have already indexed their development… (P2)

Adam explicitly brought himself to the fore in P2 and P3, but by using "the author" throughout the papers, instead of other pronouns such as "I" or "we", it seems that he tried to minimize the effects of explicitly presenting the image of the author. Through finding a middle ground through using "the author", probably more objective and author-obscure than other self-referenced pronouns, Adam was constructed as a writer who assumed some agency in his papers but simultaneously was cautious in letting out his own voice.

4.3.1.4 Engagement markers and identity construction in Adam's papers

In general, engagement markers were used to a significantly lesser extent, and only three types of engagement markers were found in Adam's papers. *Questions* were the only category that was used in the three papers, twice in each paper. Further, two types of *questions* were found in the three papers. In P1, the book review, Adam used two rhetorical questions throughout the text, whereas in P2 and P3, the research articles, *questions* were raised in the form of research questions, as shown in the following extracts:

> …and some categories would confuse readers, such as stylized Chinese Mandarin and non-standard Chinese character expressions. <u>What are differences between these two terms?</u> (P1)
>
> Two research questions were figured out: <u>1. What are the similarities and differences between C9 and international universities regarding their rhetorical moves? 2. How do semiotic recourses construct the rhetorical moves between both kinds of universities?</u> (P2)
>
> The two research questions were figured out: <u>What social identities were constructed by Michelin restaurant reviewers on RED? How did they construct the different social identities on RED?</u> (P3)

Questions underlie the essential dialogic nature of academic writing (Bakhtin, 1986). They directly involved readers' engagement with the texts and led readers through an argument (Hyland, 2002c). In P1, *questions* were brought up based on Adam's understanding of the texts he reviewed, while in the other two papers, *questions* were generated from the current body of literature based on Adam's expertise in the field. The inclusion of questions in the three papers demonstrated Adam's identity as a critical thinker who could "create anticipation, arouse interest, challenge the reader into thinking about the topic of the text" (Webber, 1994, p. 266).

Table 20. Interactional metadiscourse in Adam's papers – engagement

	P1		P2		P3		Total	
Category	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words	Token (%)	Per 1,000 words
Reader pronouns	1 (1.2%)	0.6	0 (0%)	0	3 (0.9%)	0.5	4 (0.5%)	0.2
Directives	0 (0%)	0	7 (2.2%)	1	5 (1.4%)	0.7	12 (1.6%)	0.8
Questions	2 (2.4%)	1.2	2 (0.6%)	0.2	2 (0.6%)	0.3	6 (0.8%)	0.4
Appeals to shared knowledge	0 (0%)	0	0 (0%)	0	0 (0%)	0	0 (0%)	0
Personal asides	0 (0%)	0	0 (0%)	0	0 (0%)	0	0 (0%)	0
Interactional-Engagement (Total)	3 (3.6%)	1.8	9 (2.8%)	1.2	10 (2.9%)	1.5	22 (2.9%)	1.4

Directives were only found in the two research articles, appearing 7 times and 5 times, which covered 2.2% and 1.4% of the total metadiscourse markers respectively. Though the frequencies were not high, from Adam's data, all three types of directives attaining different functions (textual acts, physical acts, and cognitive acts) were found. In the example below, Adam used imperative sentences in parentheses to direct readers' attention to other parts of the paper. To fulfill the textual acts function, *directives* were used in places where information provided could possibly scaffold readers' understandings of the text:

UIC mentions that it was the first university cooperated by Hong Kong and mainland universities (see Excerpt 2) and NYU-shanghai emphasized the first Sino-American university (see Excerpt 3). (P2)

Further, imperative sentences were found in the main texts, attaining the cognitive acts function. As shown in the following example, by clarifying what the scope of an important notion entailed in the paper, Adam initiated readers into a new perception of a concept in the way that he deemed relevant:

Note that discourse strategies in the current paper refer to social practices, including textual strategies and other semiotic resources as well (van Leeuwen, 2008). (P3)

Besides the examples above that used directives to instruct readers how they should participate in the text, in Adam's P3, he used a modal of obligation to specify the need for potential readers (i.e., researchers from the field) to perform some action outside the text (Hyland, 2002b), which was the physical acts of *directives*:

In terms of the limitation, online discourse studies should shift from language to internet users per se (Androutsopoulos, 2008). (P3)

The next category of engagement marker that was found in Adam's papers was *reader pronouns*, which only appeared in P1 and P3, once in the former, and three times in the latter. Adam exclusively used inclusive pronouns (we and our) in the two papers, as the following examples show:

As social media permeates every aspect of life and has become a necessary component of our socialization, people already entered post-digital era (Lyons et al., 2021). (P1)

… the distinction and class inequality do not disappear, but are naturalized in the diverse cultural choices including the food we eat. (P3)

RED is also famous as an image-sharing platform, or we can say it provides rich multimodal resources. (P3)

Reader pronouns are important indicators demonstrating writer's solidarity with the readers, engaging readers in the shared intellectual journey. Adam signaled the disciplinary insider identity through the employment of *reader pronouns*. Through the use of *reader pronouns*, Adam specifically demonstrated his alignment with the reader, who was the course instructor.

Appeals to shared knowledge and *personal asides*, notably, were not found in Adam's three papers.

4.3.2 Patterns of identity construction through metadiscourse use in Adam's three papers

In this section, based on the results of last section, I compare the ways metadiscourse was used to construct identities in Adam's three papers. I start with the overall pattern of the metadiscourse markers, followed by interactive metadiscourse markers and interactional metadiscourse.

4.3.2.1 Overall patterns of metadiscourse usage in Adam's papers

Overall, metadiscourse markers were used at a similar frequency in Adam's three papers, with P3 having more metadiscourse items than the other two papers. In terms of interactive metadiscourse, it was found that they appeared more frequently in P3 (43.7 times per 1,000 words) than P1 and P2 where the frequencies were very close (36.9 times vs. 35.8 times per 1,000 words). This demonstrated that in P3, Adam's identity as a writer who used explicit textual signals to glue texts together was more conspicuous than in P1 and P2.

P1, the book review, was found to have more stance markers (11.3 times per 1,000 words) than P2 and P3. In the two research papers, stance markers appeared at an almost identical frequency (7.2 times vs. 7.3 times after normalization). The difference could show that in the book review, Adam was positioned as a writer who exhibited more evaluations and personal judgment than in the research articles.

Lastly, engagement markers were used minimally across the three papers, which probably demonstrated that Adam, as a writer, did not engage his audience into his argumentation in the paper extensively.

4.3.2.2 Patterns of interactive metadiscourse in identity construction in Adam's papers

Interactive metsdiscourse markers were the most frequently used category of metadiscourse in all of Adam's three papers. A noticeable pattern emergent with regard to its sub-categories was that *endophoric markers* were not used in P1, but they constituted the same frequency in P2 and P3 (both were 10.3% of the total metadiscourse markers). This probably demonstrated that only in research papers Adam needed to refer to other parts of his texts, while in book review, a line of argument could be established without referring internally to the text itself.

Another category that was used at an almost identical frequency was *transitions*: They covered 22.7% of the total metadiscourse markers in P2, 23.1% in P3, while significantly more (40.5%) were found in P1. This distinction showed that in book review, Adam organized his text using more explicit textual signals than in research papers.

Evidentials, in a similar vein, were used almost equally frequently in P2 and P3 (11.3 times and 14.3 times per 1,000 words), but they were used significantly less in P1, the book review (3.6 times after normalization). This pattern suggested that in research papers, Adam constructed the identity of a disciplinary insider, while in book review, the rhetorical purposes of this genre probably required him to construct such an identity to a lesser extent.

Code glosses were relatively evenly distributed across Adam's three papers (8.3, 8.5 and 10.9 times after normalization). This meant that in regard to making reformulations and exemplifications to explicate the propositional entities, Adam constructed a writer identity similarly across different written genres who provided the "propositional embellishment" (Hyland, 2007, p. 267) to enhance readers' understanding and shape meanings more accurately.

Lastly, *frame markers* were used more frequently in P1 (4.8 times after normalization), and a remarkably less amount was found in the two research papers (1.4 and 0.9 times per 1,000 words). The differences could demonstrate that in the book review, Adam tended to employ more explicit textual markers to label structure of the text, while research papers had a relatively fixed rhetorical structure for writers to follow, which probably placed less emphasis on the writer to construct an identity that guided reader throughout the text using signposts.

4.3.2.3 Patterns of stance markers in identity construction in Adam's papers

Stance markers constituted a frequently used category of metadiscourse in Adam's papers. Among the three papers, they were used the most frequently in P1, the book review, and relatively less were used in P2 and P3. This probably demonstrated that Adam positioned his identity in a more conspicuous way in relation to the arguments and propositions in writing the book review than in the research papers.

Specifically, *hedges* occurred in the two research papers at roughly the same frequency (4.3 times vs. 4.5 times per 1,000 words), but they were used less in P1. Interestingly, *boosters* were also used at the same frequency in P2 and P3 (both 1.8 times per 1,000 words); however, they were found to be used much more in P1. The similarities in expressing assertiveness and uncertainty in the two research papers, and the differences exhibited in the book review probably showed that the strategies Adam adopted in maneuvering the rhetorical effects were largely related to the genre he performed: He tended to construct his identities of a cautious writer and a confident writer in the same way in both research articles, while in the book review, Adam tended to boost his tone of certainty more than in the research papers, constructing an identity of an writer who knew well of the content knowledge in the study he reviewed and who was confident in making evaluations and critiques on the study.

On the other hand, *self-mentions* in Adam's papers also demonstrated genre-specific patterns in Adam's papers: They were only used in the research papers (P2 and P3), but were not found in the book review (P1). This pattern suggests that Adam's self-representation was more prominent in the research papers, where he needed to specify research procedures using "the author". In book review, this rhetorical move was apparently unnecessary, hence his role of the author was downplayed.

Lastly, *attitude markers* were found in all of Adam's three papers, but in P1, they occurred the most frequently. This difference is mostly likely related to the evaluative purposes of book reviews, which required Adam to critique the study he reviewed. Hence in the book review, Adam constructed an identity of an evaluator and critic more prominently than in the research articles.

4.3.2.4 Patterns of engagement markers in identity construction in Adam's papers

Engagement markers were not a category prominently used in Adam's three papers. Among them, *appeals to shared knowledge* and *personal asides* were not found in the papers. *Questions* were found in all three papers, with different types of *questions* being used. *Directives*, however, were only found in the two research papers. This showed that in the research article genre, Adam may need to demonstrate himself as an authoritative writer who directed readers' acts of different kinds (such as cognitive, textual and physical acts reported in section 4.3.1.3), whereas in the book review, it was less prominent to instruct readers to see things or perform actions determined by Adam but to present his critiques and stances toward the study under review. Lastly, *reader pronouns* were used in P1 and P3.

4.3.3 Factors shaping Adam's identity construction in the three papers

In this section, I report factors that influenced Adam's identity construction embodied through metadiscoursal resources in his three papers.

4.3.3.1 *Autobiographical self* and *discoursal self* and *self as author*

4.3.3.1.1 Previous academic trainings

As stated in the methodology chapter, Adam took linguistics related courses in his undergraduate program. He had taken Corpus Linguistics, and from the course he read a paper which was about *self-mentions*. He learned that the paper arrived at a conclusion that only established scholars in the field could use self-referenced pronouns. As he said,

> Adam: Yeah, I feel… because I had taken a course named *Corpus Linguistics*. A paper we read from the course was about self-mentions. It talked about only big names can use first person pronouns.
>
> Ren: So you think now you're a newbie, you are not supposed to use self-referenced pronouns.
>
> Adam: Yes, I do have this feeling. I feel that's too subjective. (I2)

In the meantime, Adam recollected the profound influence of undergraduate courses on his subsequent academic writing: "I don't prefer to use "I" in academic papers… What I learned in my undergrad program had deeply rooted influence." (I2)

4.3.3.1.2 Being a fresh PhD student

Adam's feeling of himself not being supposed to use self-referenced pronouns was clearly exhibited in his book review (P1), where no cases of *self-mentions* were found. As mentioned in the methodology chapter, before writing the book review, Adam was required to write a paper comparing three book reviews written by well-known scholars. His way of thinking that he should not use *self-mentions* was reflected and substantiated in the comparison paper, in which he wrote:

> "For instance, the author added her suggestions when she reviewed how to find a potential journal. 'I would add one criterion: speed of publication, as researchers just starting to publish in particular find it important get into press as soon as they can' (p. 52). This critique and subjective style might result from the status of the reviewer, who has been experienced editor of academic journals."

By highlighting the academic status of the author (who was an "experienced editor of academic journals"), Adam implied the "inappropriateness" that his *autobiographical self* as a fresh PhD student brought to him in regard to the use of *self-mentions*. Further, he particularly specified his reasons for not using self-referenced pronouns in our interview:

> I feel my opinions are too subjective. You're not big names in the academia… Or I normally use "the author", I don't use "I". I remember when I wrote my undergraduate paper, we discussed that you cannot achieve objectivity in humanities or social sciences. Those disciplines have innate biases. (I1)

4.3.3.2 *Possibilities for selfhood* and *discoursal self* and *self as author*

4.3.3.2.1 Target audience

As stated in the methodology chapter, Adam's papers were from two courses which were taught by the same instructor. Since Adam's papers were written to complete course assignments, when writing the papers, he had clear audience awareness that the target audience of his papers was the course instructor. The course instructor's research interest and the studies the instructor had conducted had influenced Adam's topic selection. As he said in our interview for the book review:

Ren: Your instructor's interest, is that related to your choice (of this study)?

Adam: Yes, yes.

Ren: You want to be close to your instructor's research area.

Adam: Yes, I want to work with my instructor. (I1)

More specifically, Adam's audience awareness shaped his *evidentials* use in the paper. For instance, in P2, Adam consciously cited articles published by the course instructor, as he said:

> In that paper, I cited a lot of papers of my instructor's, because I collected data from XXX (the social media platform), about food. My instructor also published a paper about that. (I2)

Target audience did not simply affect Adam's *evidentials* use in P2; in writing P3, he acknowledged the influence of the course instructor, who was the only audience he was thinking of when writing the paper. Adam said he read and cited a lot of the course instructor's papers:

Adam: I read a lot of her papers, her previous papers.

Ren: I also noticed you cited a lot of the course instructor's papers. Do you think if this is related to the fact that she is your reader? You're my reader, so I want to show you I've read your work.

Adam: Yes, I agree.

Ren: Yeah, and also she is your grader.

Adam: Yes. (I3)

4.3.3.2.2 Instructor's feedback

Instructor's feedback was important in shaping Adam's identities, which was detail-oriented and focused on various categories of metadiscourse usage. For example, the instructor made suggestions on the *evidentials* used in Adam's P3 by not only commenting positively on the usage, but also making recommendations on sources Adam could have referred to. The instructor specifically provided a piece of work that could potentially be helpful with Adam's understanding of the topic in question.

Further, the instructor made concrete suggestions regarding the presentation of the literature review in Adam's P3. Adam's way of organizing the literature seemed inconsistent, which raised the instructor's concern. Therefore, the instructor highlighted a string of citations and commented that the list should be revised, and the literature review should be reorganized accordingly. The instructor's feedback on *evidentials* use in Adam's P3 would facilitate Adam's identity as a competent writer who could present relevant studies in a systematic way.

Moreover, the instructor specifically recognized the *frame markers* used in Adam's paper. The instructor highlighted the use of the *frame marker* (In a nutshell) that labeled stages in the following sentence:

> In a nutshell, the second section discussed the digitalized food and reviewers' identities… (P3)

The comment went:

> I appreciate that you're using signposting to guide your reader!

Interestingly, though Adam personally tended to avoid the use of *self-mentions*, using the self-referenced pronouns seemed to be advocated by his instructor. In the methodology section that described data collection process in P3, Adam used "the author" to refer to himself. Commenting on this usage, the instructor made a suggestion on Adam's habit of avoiding using *self-mentions*:

> I strongly recommend getting used to writing in the 1st person. YOU are the one making these sampling decisions, so there's no point in trying to hide that.

4.3.3.2.3 Writing conventions

Adam remarked that his writing was influenced by the established writing conventions in the field. In other words, he imitated writing style from published articles, which he perceived to be acceptable in the field. Specifically, he appropriated the usage of *attitude markers* and applied them in his own paper (e.g., "it is surprising that…" found in his P2). He explained the reasons for writing this way:

> For this, I saw other people write this way in the literature, so I also write this way… I think it aligns with the conventions… and since those papers were published in journals, it means that's an acceptable convention. I think it's a safety space. So here I can write this way. (I2)

Field conventions did not only influence Adam's use of *attitude markers*. In his P3, he wrote a sentence starting with "Note that discourse strategies in the current paper…". He stated that including the *directives* in his paper was due to the writing style he saw from his previous readings. As our conversation went:

> Adam: This is because, I saw a big name in this field used the structure like this, so I also used in in my paper.
>
> Ren: Did you realize the imperative sentence like this has some tone of commanding?
>
> Adam: I did. But I feel if other people use it, I can too. (I3)

Conventions in the field preferred by established scholars also influenced Adam's choice of *self-mentions*. For instance, in his book review, he did not use self-referenced pronouns. This was partially due to his real-world identity as a PhD student, as reported above, and partially because of the conventions he observed from the book reviews he read. As he wrote in the comparison paper on the three book reviews:

> "In terms of the language style, the review from (Author #1) is most 'objective' among these reviews, and the (Author #2)'s BR is more objective than that of (Author #3). (Author #1) seldom used the pronoun 'I' in her paper, and most of the subjects in this review are inanimate subjects, such as 'chapter', 'section' and 'book'."

Noticing that the three authors of the book reviews on which he based his own book review were "objective" as they barely used self-referenced pronouns, he adopted a similar writing style and avoided the use of *self-mentions* in his book review.

However, since Adam's book review was heavily based on the three book reviews he analyzed and consulted, the influence of established scholars' writing was not reflected only through his way of presenting himself through avoidance of *self-mentions*. When I asked him his opinions in regard to the use of some adjectival phrases (*attitude markers*) that expressed his affective stance, he remarked those were imitations of the three book reviews. As he said:

> Ren: Like here you have some evaluative expressions, like "it is helpful for to refer to his collection strategy…"
>
> Adam: That's an imitation of the three book reviews. (I1)

In writing his own book reviews, Adam followed the structure of book reviews he learned from the papers he read. So, he had the knowledge that adding evaluations was a required move in writing book reviews. Through using *attitude markers* in the book review, he was aware that expressing attitudinal stance could help construct his identity as a professional writer, as he said:

> This (using adjectives to express your attitude) will show that you're professional. It means you can notice the… you're critiquing this kind of… And book review has its structure. One paper about how to write book reviews, and it says in the end, you have to add your own evaluation. (I1)

4.3.3.3 Agency and *discoursal self* and *self as author*

Adam seemed to be the writer who demonstrated his awareness of and purposefulness in constructing his identities using certain metadiscoursal resources. When using *directives* to highlight the information Adam wanted to clarify for his readers in P3 (The sentence went: Note that discourse strategies in the current paper…), he stated using this metadiscoursal item could reflect his identity as an author of this article who had an agentive position in directing readers. As he said:

> Because I am the author of this paper, I have the authorship, so I can frame my argument (the way I prefer). Well, this is only how I frame my argument. (I3)

Adam also used his knowledge of relevant methodology and of the field to frame his argument through various kinds of metadiscoursal markers such as *transitions, frame markers* and *attitude markers*, as shown in the following two examples extracted from the book review:

> <u>Secondly</u>, <u>though</u> the author presented the comprehensive backgrounds of two platforms, he <u>still</u> did not rationalize why he chose these two sites and why to compare them. (P1)
>
> <u>Although</u> the author stated that paper was to investigate a generalized and regularity pattern of online creative literacy in Chinese context, <u>it is better to</u> interview some users to know their emic views and ideologies for certain semiotics. (P1)
>
> The author also overlooked the 'interactional' nature of social media, <u>though</u> he stated it was an important characteristic of social media. (P1)

Through using the metadiscoursal items, Adam pointed out the significance and contributions of the study he reviewed, but he also took a further step to highlight the flaws that he considered still existed based on his expertise of the field. Showing his being a critical writer, he agentively presented his critical arguments using the metadiscourse markers. His logic of reasoning went:

> I do always think that it does has the flaws, because they (methods and analysis) were not related. Because since the author said he used XX (a certain methodology), he cannot just analyze the text. He should contact the person. Yeah, so that's my understanding, based on my understanding of XX (the methodology). (I1)

4.4 Comparison of the three cases: identity construction

In this section, based on the results of the three cases reported in the previous parts of this chapter, I make a comparison among the cases and present the similarities and differences with regard to their identities constructed in the papers through the employment of metadiscourse as well as factors contributing to their identity construction.

The general patterns of metadiscourse use were similar given the fact that interactive metadiscourse was the category that was overwhelmingly used in all the participants' papers. Further, stance markers were used to a lesser degree, while engagement markers appeared negligibly in all the participants' papers. As claimed in a previous study, "identities are patterned" (Kamler & Thompson, 2014, p. 19), which could reflect the norms and frameworks available in the community. While the finding that the majority of metadiscourse markers used in the participants' papers was interactive metadiscourse seems to be in line with that of some of the previous studies examining graduate students' papers (Lee & Casal, 2014) and research articles (Hyland & Jiang, 2022), the same pattern demonstrated in the three cases in this dissertation study probably shows that the three participants, regardless of their disciplinary background or the nature of the writing task, attached significantly more importance to the internal coherence of their papers through using explicit linguistic signals. Organizing texts in a logically coherent way contributed to the construction of their identities as competent academic writers.

However, though the similarities were prominent in the participants' papers, disciplinary differences are expected to occur in writing, because the rules change from discipline to discipline, and the audience and purposes of writing also vary across contexts (Paltridge, 2004). Despite the fact that the papers written by the three participants demonstrated identical patterns with regard to the categories of metadiscourse, Adam's papers were found to have more metadiscourse markers than Momo's and Nancy's papers. This seems to support previous findings (e.g., Hyland, 2000, 2018) that in soft disciplines, writers tended to employ more metadiscourse markers. This showcases the differences in identity construction via metadiscourse resources that could be accounted for by disciplinary differences: In soft disciplines, which are more discursive and require extensive verbal reasoning and argumentation (Hyland, 2012a), more metadiscourse would be used to reflect the discursively elaborate nature of the written discourse from the disciplines. However, in hard disciplines, where reasoning is more constructed on objective experimentation and calculation, fewer metadicourse markers denoting textual relationships and reader-writer relationships is largely a preferred pattern, as pointed out by Hyland (2002b), "While all writing needs to solicit reader collusion and carry an audience through an argument, the most explicit interactional signals tend to cluster in humanities and social science papers" (p. 537).

Engagement markers, compared to interactive metadiscourse and stance markers, were used negligibly in all three participants' papers. Specifically, *questions* were not used in Momo's and Nancy's papers except in Nancy's dissertation where she needed to put forward the research questions in a separate section. This may have to do with the lengthy nature of dissertations; specifying research questions explicitly is convenient for readers to follow the information. Lack of *questions* seems a common phenomenon in the academic writing in the hard disciplines. Hyland (2018) pointed out that *questions* are mainly confined to the soft fields. Further, Hyland's informants from the hard fields specifically used "intrusive" and "condescending" (p. 187) to describe the inclusion of *questions* in academic articles. However, *questions* were found in all Adam three papers. In soft disciplines employing *questions* is seen as an important strategy in advancing dialogues and involve readers in the text. This way, the participants' identities from the hard discipline in this study were constructed as writers who reported the research and required readers' involvement to a lesser degree, but for soft discipline writers, *questions* helped

Adam build an authoritative identity who invited readers to a frame of discourse and simultaneously led readers to his viewpoint (Hyland, 2002c; Hyland & Jiang, 2016).

Writers were not necessarily consistent in constructing their identities across texts. Specifically, this study found the participants varies in conveying authoritative identities using self-referenced pronouns in their writing (Ivanič, 1998). For example, Momo did not use *self-mentions* in P1, the research article she wrote in the first semester of her PhD; Nancy's P2, the government report, did not employ *self-mentions*. This may demonstrate the inconsistency of writer identities constructed by the same writer. Further, the results indicated that an inconsistent representation of the writers' identity within a piece of paper might emerge. For example, *hedges* and *boosters* were found to be the most frequently used interactional metadiscourse markers almost across all the papers, and they conveyed the opposite connotations in positioning the writers' stance. As identity is multiple and fluid, it is important to note that identity construction is not necessarily uniform or coherent. Even within one single piece of text, it may be common that writers shift from one subject identity to another, which could then create multiple and sometimes contradictory impressions of themselves (Clark & Ivanič, 1997). A constellation of multiple (even conflicting) identities could reside in one single piece of text.

Possibilities for selfhood played an indispensable role in providing linguistic repertoire available for writers to construct their identities. The essential role of various contextual variables in shaping the construction of identity has been substantiated in previous research (e.g., Chiu, 2016; Hirvela & Belcher, 2001). With the available pool of linguistic resources, writers have sufficient field knowledge and rhetorical knowledge to frame their papers in order to write in a style accepted and preferred by the discourse community. *Possibilities for selfhood* are vital in constructing the three participants' identities. As demonstrated by the results, field conventions are a significant factor that shaped writers' identity construction. They included the writing preferences by established scholars, as shown in Adam's and Nancy's case, from which the two writers imitated the metadiscoursal usage, such as *attitude markers* and *frame markers*, to construct their identities. In Momo's case, she used *directives* to point out future directions for practice and research since this was a common rule preferred by the field. These metadiscoursal features represented the dominant interests, knowledge making and knowledge telling practices in their respective discourse communities (Clark & Ivanič, 1997), and the participants accepted and obeyed the prevailing norms that were valid for the identities they constructed or aimed to construct (Weidman & Stein, 2003). Learning from the established community practices was a key means for acquiring the discourse, as "communities provide the context within which we learn to communicate and to interpret each other's talk, gradually acquiring the specialised discourse competencies to participate as members" (Hyland, 2015, p. 33).

Another conspicuous strand of *possibilities for selfhood* emergent from the study was guidance from senior researchers. Advisors' feedback is essential for students to make improvements on their academic writing skills and disciplinary knowledge expertise (Xu, 2017) and facilitates their identity construction as credible and competent scholars (Li & Han, 2023). This influencing factor that shaped writer identity was found in all three participants' cases in this dissertation study. The important role of academics from the participants' institutions in the shaping of the writing and identity could find support in Botelho de Magalhães's (2019) study where two L2 doctoral students interacted with scholars from their university extensively, mainly their principal supervisors. In Wang and Parr's (2021) study, the PhD student was also found to engage in dialogic activities with her supervisors extensively, which contributed to her significant growth as a writer. The participant, who was an L2 PhD student studying in Australia in Xu and Zhang's (2019) study, demonstrated that how the student's interactions with the discourse communities facilitated her identity growth as a researcher holding different epistemologies. Another important factor under *possibilities for selfhood* that mattered significantly in the doctoral students' identity construction was comments from journal reviewers. This was obvious in Momo's and Nancy's cases since the data they shared with me included research articles submitted for publications. Reviewers' feedback contributed to their identity construction and helped them present identities of scholarly writers.

In regard to *autobiographical self* and the identities constructed through papers, it was found that the participants were largely influenced by the academic training they had in China. Before the participants started their PhD programs in the U.S., they all brought linguistic and academic capital to their doctoral writing. It was found that the case in Wang and Parr's (2021) study applied her previous academic writing experience in her home country (China) and the foreign country where she subsequently studied to the academic writing during her PhD. Specifically, the skills of identifying and fixing grammatical and lexical errors had boosted her confidence, which significantly impacted her writing and identity in her doctoral studies. Further, the experience of studying in a foreign country helped her develop some preliminary agency in her English writing. Xu (2017) reflected on her own experience as an L2 doctoral student studying in an ESL setting, and concluded that her life history, such as the ideology she was immersed in in her home country served as a departure for her to co-construct the supervisory relationship, which affected her ways of dealing with advisor's feedback.

From a social constructivist perspective, identity is constructed collaboratively by sociocultural context and individuals' agency (Matsuda, 2015). This said, in this dissertation study, agency seemed to play a relatively minor role compared to *autobiographical self* and *possibilities for selfhood* that influenced the three writers' identity construction particularly in the two cases from the hard disciplines. The writers only exerted their agency within the frame of limited available options. That is to say, within the frame of disciplinary or academic discourse, linguistic resources writers could choose were limited to construct their identities (Ivanič, 1998). For example, Momo and Nancy did choose from the available linguistic repertoire to construct their intended identities based on the nature of content (whether the propositional information was their interpretation or objective results that were achieved by experimentation); however, they were not conscious of some metadiscoursal choices they made that could potentially construct their identities to readers.

CHAPTER 5: DISCUSSION AND CONCLUSION

In this chapter, I first discuss the findings of each case in relation to the identity scholarship in the current field. Then I present the insights generated from the current study that could potentially inform doctoral academic writing and identity construction. Lastly, I point out the limitations existing in this dissertation research, and suggest recommendations for future research.

5.1 Discussion on Momo's case

Momo's three papers demonstrated that the distribution of metadiscourse largely followed that same pattern: Interactive metadiscourse was used the most frequently, while stance markers were found to appear to a much lesser extent, and engagement markers were used negligibly. Though the distribution of each category of metadiscourse markers was the same in the three papers, in P3, the paper aiming for publication, metadiscourse markers were found to be used more frequently than P1, a paper for publication, and P2, a paper for conference presentation.

The linguistically more sophisticated language use in P3 than P1 and P2 due to the more frequent usage of metadiscourse markers indicated that the construction of Momo's identities varied over time and across writing of different genres. This difference can probably be accounted for in the following ways. First, to look at P2 and P3, the two versions of a paper of the same topic written for different purposes, more metadiscourse markers were used in P3. To break it down into the sub-categories of metadiscourse, each category of metadiscourse was used more frequently in P3 than in P2. This possibly showed the genres Momo performed could have influenced the metadiscourse use, as Momo was aware that P2 would not get published, and the purpose was mainly to present the study at the conference. Hence the ideational purpose (Halliday et al., 2004) of the paper, which was to present the content of the study, possibly overweighed the interpersonal and textual purposes of the paper, which were to create a logically and textually coherent paper in order to achieve academic persuasion among the target readers. The purposes of knowledge making affected the metadiscourse use in the two papers, which subsequently shaped Momo's identities constructed in the two papers. On the other hand, the difference in regard to metadiscourse use between the two versions of the same paper could be due to the multiple rounds of revisions after Momo attended the conference and then aimed to get P3 published. As she said in our interviews, senior researchers from her field, who were the conference abstract reviewers and the co-authors of P3, had offered feedback covering various aspects of metadiscourse use, which could have facilitated her revision and established her identity as a scholarly and professional writer.

Second, to compare P1 and P3, the two papers written for journal publications, P3 used more metadiscoursal markers than P1, and each metadiscourse category demonstrated more occurrences in P3 than P1. As has been reported in the methodology chapter, P1 was written in the first semester of Momo's PhD and P3 in the beginning of her third semester. Therefore, the changes could be due to the apprenticeship Momo went through in the first year of her PhD. This changing pattern supports the argument that identity is not a fixed construct; instead, it is always in a flux, which constantly undergoes changes under various academic and/or rhetorical contexts (Ivanič, 1998).

To look at each category of the metadiscourse markers, a distinctive pattern could possibly reflect the above discussed pattern. That is, each category of interactive metadiscourse in P1 was found to occur more frequently than in P2, and in P3, interactive metadiscourse was found to occur more than P1 and P2. This again confirmed that the rhetorical purposes of P2 perceived by Momo were to present the information; to organize the text coherently using interactive metadiscourse was a secondary goal for this paper. However, fewer occurrences of interactive metadiscourse in P1 than in P3 probably show that Momo as a writer had accumulated textual awareness over time, which could also be attributed to the help and guidance of senior researchers in her field and the co-authors. Thus, Momo constructed her identities diachronically as a writer who could gradually manipulate textual resources more skillfully to attain the purposes of her papers.

The use of stance markers, however, displayed more complex patterns than interactive metadiscourse did. A noticeable distinction was that in P1, no *self-mentions* were found, but they were both used in P2 and P3. The developmental pattern probably reflected that Momo's perceptions of this marker conspicuously denoting the writer's identity (Tang & John, 1999) changed overtime, as she pointed out in our interview that the writing conventions in the top-tiered journals changed her attitude toward this metadiscoursal resource. Note that Momo's P1 was a collaboration with her advisor from her master's program in China; therefore, a change of academic environment when writing P2 and P3 together with the co-authors she was working with probably could provide partial explanation for this transition regarding the use of *self-mentions*.

Momo's three papers were co-authored with her advisors or senior researchers, whose roles in conceptualizing and revising the papers to facilitate Momo's identity construction as a professional and scholarly writer were reported in the results chapter. As a common practice in hard disciplines, "co-authorship with supervisors is a significant pedagogic practice that can enhance the robustness and know-how of emergent scholars" (Kamler, 2008, p. 283). It is true that Momo benefited greatly from the collaboration with the senior researchers, who were an invaluable source from the community of practice that could directly contribute to Momo's identity construction in academic papers.

Previous studies have demonstrated that there may be discrepancy regarding identities constructed by the writer and those perceived by readers (Matsuda, 2015). Momo did not realize the differences of some *hedges* and *boosters* (such as "demonstrate", "indicate", "show", etc.) – she thought they were the same in connotations. The lack of awareness of the usage of some metadiscourse markers indicates that the identities constructed in her papers in relation to some *hedges* and *boosters* were not intended – There's inconsistency between the identities writers intend to project and the ones that are actually perceived by readers (Matsuda & Tardy, 2007). Momo's not realizing the identity projected through the use of some metadiscoursal markers also showed that "adopting the practices in which they are embedded… is almost always subconscious, as are most processes to do with ideologies" (Clark & Ivanič, 1997, p. 139).

Further, the results showed that Momo exerted agency in order to avoid repetitions regarding the use of *self-mentions* in her papers through substituting the pronouns with "this paper", "this project" and so on. This could demonstrate that Momo did not fully understand the functions of those choices. That is to say, her understandings of *self-mentions* in academic writing were partial, based on the literature she read. She was not equipped with the knowledge that what those words could bring to the construction of herself as a writer in her papers. Lacking explicit knowledge in this regard probably shows that Momo's identity construction is an unconscious process, which was shaped by the contextual variables during her socialization into the discourse communities.

To conclude, Momo's case showed that identity construction varied across different genres as well as over time; however, that she was not aware of the functions of her metadiscoursal choices seemed to project identities that might be perceived differently by the readers.

5.2 Discussion on Nancy's case

Nancy performed different genres in the three papers examined for this dissertation research. They included a journal article, a report for government, and her dissertation. Metadiscourse markers appeared at a rather similar frequency across Nancy's three papers: interactive metadiscourse was used more than 70%, stance markers more than 20%, while engagement markers only showed up less than 2% in the three papers. However, each type of metadiscourse was found to be used the most frequently in P3, the dissertation. This may have to do with the genres Nancy performed in these three writing tasks. Given the fact that dissertation was a lengthy project, Nancy may need to make greater use of metadiscourse markers (particularly interactive metadiscourse) to structure the discursively complicated arguments (Hyland, 2018). Meanwhile, the higher frequency of stance and engagement markers in the dissertation also seemed to demonstrate that Nancy as a writer made more sophisticated efforts to engage with her readers and present herself as credible student writer and researcher informed by the field knowledge and practices (Hyland, 2018).

Though Nancy's P2, the report written for government, was a relatively lengthy project as well, fewer metadiscourse markers were found in this genre. This could represent that there were substantial variations in regard to the use of metadiscourse in writing for different purposes in Nancy's case. In writing the report, there was a template Nancy had to follow strictly, and she remarked that the report was a monotonous genre to perform: All she had to do was to present and describe the data collected. This had largely mitigated the necessity of making use of metadiscourse to structure the texts and engage with readers in the report. This difference in terms of metadiscourse use positioned Nancy's identities differently in the two academic genres: One mainly was a presenter of information, overshadowing the need to represent oneself as a critical and competent academic writer, while in the dissertation, a scholarly identity well-informed of the field knowledge was expected.

Specifically, in terms of interactive metadiscourse, the *evidentials* used in the three papers seemed also to reflect the influence of genre on identity construction with regard to citation use. *Evidentials* were used significantly less in P2 than in P1 and P3. This demonstrated that in the report, to pool a vast amount of literature to support the study was a less prominent focus compared to the journal article (P1) and dissertation (P3), where Nancy had to present her identity as a credible writer who had read extensively relevant up-to-date research in order to write persuasively. *Endophoric markers*, on the other hand, were used significantly more in the report than in P1 and P3. This again showed that in report writing, more information was presented in a multimodal way, which required Nancy to refer to those Figures or Tables using extensively *endophoric markers.* By doing this, she constructed herself more effectively as a presenter of the information gathered by the research team, and the information could be approached by the readers more effectively.

Interactional metadiscourse was relatively evenly distributed in Nancy's three papers. It is worth our attention that though *self-mentions* were found in P1 and P3, they were not used in

P2, the government report. This demonstrated the differences in regard to identity construction in relation to genres. It implied that Nancy's identity as the (co-)author of the report was downplayed, and she (and the research team) did not assume the agentive role of the paper. In the report, Nancy did not have to follow the pattern of research articles to represent her identities through *self-mentions*. She only was required to present the important information to the project manager/s in a procedural way following the preferred conventions prescribed in the template.

Similar to Momo's case, not realizing that linguistic choices could construct identities was emergent in Nancy's case. Specifically, Nancy's choice of some metadiscourse markers was guided not by the explicit knowledge of those linguistic resources, but mostly by her intuition. As demonstrated in the results chapter, Nancy used *attitude markers* such as "surprising" or "interesting" in her P2, the government report, but she did not seriously consider the impact of "surprising" or "interesting". Her use of those words that could express affective stance was simply for the purposes of promoting the diversity of language use in a genre which she considered "monotonous", without taking into consideration how using the adjectives would project images of herself on the readers. Further, Nancy used *hedges* such as "may" just to avoid making grammatical mistakes. Avoiding using *self-mentions* and using passive voice instead was just her writing habit. All these results indicated that she did not realize the functions of those metadiscourse markers, which could help her construct identities through her papers. Not consciously realizing how certain words could project her persona as a writer through text, Nancy probably constructed identities that were not her initial intentions. This also highlights the issue that Nancy seemed to lack the critical knowledge of the functions of language and metadiscourse in constructing her identities through her writing. That is to say, words do not only perform ideational and textual functions (Halliday & Matthiessen, 2014); they simultaneously could project her images to readers, which could then help strengthen or sabotage her academic credibility as a writer. Meanwhile, selecting certain metadiscourse markers not for the purpose of constructing an intended identity, Nancy mostly likely did not anticipate readers' reactions to her papers, and the identities perceived by readers probably were inconsistent with the ones she projected.

In terms of the use of *self-mentions*, particularly "we", Nancy stated that this was due to her writing habits, and she seemed to not have realized how *self-mentions* could project her identities. Further, Nancy's writing habits varied in writing for different genres. For example, per her own explanations, using "the research team" and passive voice in P2, and "we" in P3 were all due to her habits. This may be tricky to explain, but I attempt to interpret her habits as her subconscious writing preference informed by the broader writing conventions, that is, *possibilities for selfhood*. In other words, during the process of being socialized into the target discourse community, Nancy developed her idiosyncratic ways of self-representation through *self-mentions* in different genres. Though she attributed her choice to her habits, she was subconsciously able to "select" corresponding words to represent her identities in various genres. In writing the report for the state government, the writers were a fixed group of researchers, and using "the research team" or passive voice may be more appropriate to construct an objective identity, which can augment the persuasiveness of the information the research team present to their sponsors, who were not necessarily members of Nancy's academic community. However, in the research articles, Nancy's habit was to use "we" instead of a specific term referring to the co-authors. This may be related to the need of highlighting authors' roles in conducting the research and establish their credibility as academic writers.

Nancy did not consciously choose from the available repertoire to project scholarly identities, but in the meantime, she appropriated relevant metadiscourse resources that demonstrated that she was a critical and reflective writer (such as expressing her attitude through *attitude markers* and using *self-mentions* differently in different genres). Somehow, the discrepancy (and probably the conflict) between the writer's unawareness to construct identities through the linguistic choices she made and the actual manifestation of her identities probably showcased that writers' identity construction is largely a subconscious process, and identity construction is institutionalized and socialized. In other words, when writers are exposed to the academic and disciplinary environment, from which they acquire the conventions and rules preferred by the discourse communities and internalize those preferred ways of writing, their identities as a member of that community could be demonstrated whether they intend it or not (Hyland, 2002a), since effective academic writing and identity project "involves writers' selecting and deploying community-sensitive linguistic resources to represent themselves, their positions and their readers" (Hyland, 2014, p. 120).

Hyland (2015) discussed the notions of proximity and positioning in the construction of identities, with the former referring to "the relationship between the self and community" and the latter "the relationship between the speaker and what is being said" (p. 33). That is, proximity concerns norms and conventions preferred by discourse communities that writers would follow to construct their membership status, whereas positioning leaves room for flexibility so that writers' personal creativity could play out while appropriating the desired discourse (Hyland, 2012a). Though the results suggested that contextual variables played a prominent role in Nancy's construction of identities in the papers, it was obvious that Nancy as a writer tried to manipulate her positioning within the frame of community practices, as specified by Clark and

Ivanič (1997), "Individuals can only have identities that the conventions they are drawing on allow them to have" (p. 137). For example, Nancy had to strictly follow the template when writing the government report, but in the meantime, to reduce the robotic tone, she added *attitude markers* to diversify the writing style, which could reflect her identity as a reflective writer.

In a nutshell, Nancy's ways of constructing herself through text were mainly due to the influence of contextual variables (*possibilities for selfhood*) and personal experience and history (*autobiographical self*). Also, her agency was a noticeable factor contributing to the construction of her identities the way they were.

5.3 Discussion on Adam's case

As demonstrated in the methodology chapter, Adam's data consisted of two research papers and one book review. The overall distribution of each type of metadiscourse markers across the three papers displayed a similar pattern: interactive metadiscourse was used at an overwhelmingly more frequency than the other two categories of metadiscourse. This pattern is likely showing that metadiscourse usage by Adam was largely similar in his papers written for different purposes. However, the overwhelming usage of interactive metadiscourse markers in Adam's papers seems inconsistent with Hyland's (2018) study where postgraduate students' theses and dissertations were examined. His study found that in the field of applied linguistics, slightly more interactional markers (stance and engagement) were used in the postgraduate students' data. However, this difference could be accounted for by the different written genres in the two studies: Hyland's study investigated genres for graduation purposes, while this dissertation examined Adam's papers that aimed to complete course requirements. The less use of interactional metadiscourse in Adam's papers could also be explained by the target audience he had in mind when writing the papers. It can be conjectured that dissertations examined in Hyland's (2018) study aimed for a larger group of audience who may come from diverse background; in Adam's case, however, he was aware that the only reader he was writing for was his course instructor. This probably restrained the need for Adam to engage readers extensively, and the projection of an engaging writer identity in his papers. Besides, another attempting explanation for the difference may be that the graduation dissertations were completed toward the end of the postgraduate studies in Hyland's study, and those papers mostly likely had undergone rounds of revisions, whereas Adam's papers were mainly the efforts of his own. It may also has to do with the difference in the writers' awareness in the two studies. Since the students in Hyland's study were graduating, they may have developed more awareness of deployment of linguistic resources as a result of their writing experience over time whereas Adam was in his first semester.

The findings of Adam's case showed that the frequencies of some of the metadiscourse markers were almost identical in Adam's P2 and P3, including interactive metadiscourse (*endophoric markers, transitions,* and *evidentials*) and stance markers (*hedges* and *boosters*). This distinctive pattern was interesting, and it could probably demonstrate how papers of different purposes shaped Adam's identity construction through certain metadiscoursal items. As genres consist of structures and rhetorical moves that writers are expected to follow (Worden, 2018), P2 and P3, the papers written to fulfill courses' requirements as the final projects, and both being research articles genres, could exhibit the same pattern for identity construction. Note that the two papers were also written for the same instructor. Yet, the similarities in regard to identity construction in the two papers could also be explained by the time frame Adam completed the two papers: They were both completed almost at the same time toward the end of the first semester in his doctoral studies. Identity is not a fixed entity; it is fluid, and subject to time and context (Flowerdew & Wang, 2015; Ivanič, 1998); however, identity construction also involves stability. The construction of Adam's identity through some aspects of the use of metadiscourse in the two papers could demonstrate that at a given point, his identities were constructed in a largely stable and predictable way.

Another theme emergent from Adam's case was that different from Momo's and Nancy's cases, target audience greatly shaped the way he constructed his identities. For example, in his two research papers, he cited the research published by the course instructor to demonstrate his identity as a conscientious student and a disciplinary insider who shared similar research interest with the target reader. As a first year PhD student, Adam purposely aligned himself with the reader who was a member of the target discourse community to construct an identity of a member of community. Being a fresh PhD student, Adam's *autobiographical self* demonstrated that he was still in the process of gaining full membership of the target discourse community. In order to claim membership, student writers "have to appropriate (or be appropriated by) a specialized discourse, and they have to do this as though they were easily and comfortably one with their audience, as though they were members of the academy" (Bartholomae, 1986, pp. 4-5). In other words, they probably needed to "learn to speak our language" (p. 5) that they may not fully acquire yet.

Writers' purposes of their writing could determine the characteristics of their texts as well as the positions they assume in relation to the body of knowledge (Clark & Ivanič, 1997). Target audience being an important factor affecting Adam's identity construction, how readers responded to and perceived his writing was crucial for Adam. Writers' selections of textual resources to project identities "depend on whether it is necessary for them to maintain a good standing in the eyes of their readers" (Clark & Ivanič, 1997, p. 144). For Adam, the papers he wrote simultaneously fulfilled two goals: to complete course assignments as well as to project a credible student writer identity and leave good impressions on the instructor. Adam's conscious and strategic construction of his identity (such as including *evidentials* that were works of the course instructor) was motivated and mediated by his goals for his PhD studies (Cumming et al., 2002; Pavlenko & Blackledge, 2004), as he mentioned in the interview that he wanted to work with the instructor in the future.

Further, the results showed that Adam seemed to be the only participant in this study who demonstrated the awareness that identities could be constructed through his linguistic choices. For example, he was very cautious in using *self-mentions* and he implied that the reasons for avoiding using *self-mentions* was that using alternative expressions such as "this study", "this paper" could reflect more objectivity. This could find support in Hyland's (2002a) study where some of the participants claimed first person pronouns were closely related to subjectivity which was considered inappropriate for academic discourse, hence they tended to reduce attributions to self. The manifestations of self-referenced pronouns, in the meantime, indicated the extent to which writers had the sense of confidence to write with authority and to establish authorial presence (Ivanič, 1998), and "how much they feel themselves to be not just 'writers' but also 'authors' with the authority to say something" (Clark & Ivanič, 1997, p. 150). That Adam was hesitant in using self-references pronouns to signal his authorial stance was inextricably related to his *autobiographical self*, who was a fresh PhD student in the community and was still in the process of being socialized into the target discourse community.

5.4 Insights for doctoral identity construction in academic writing

This study examined identity construction through metadiscoursal resources in three L2 doctoral writers' papers, and uncovered the variations in regard to identity construction in one individual doctoral writer's academic papers of different types and the differences across disciplines. Meanwhile, this dissertation generated contextual and personal factors that had a swing on L2 doctoral writers' identity construction. Given the results, some insights may be gleaned concerning L2 doctoral identity construction in their academic writing.

As reflected in Momo's and Nancy's cases, both of the doctoral writers from the engineering fields did not realize the nature of some metadiscourse items, which could help construct corresponding identities; instead, what drove their selection of the metadiscourse markers was their intuition or the functional purposes of metadiscourse usage such as avoiding repetition or grammatical mistakes. As issues of identity are fundamental to writing (Ivanič & Camps, 2001; Ivanič, 1998), and language choices affect both meaning conveyed and the impressions of the writer perceived by readers (Ivanič & Simpson, 1992, p. 144), it is necessary to raise doctoral writers' critical language awareness (Fairclough, 2014) so that they are able to recognize the types of identities they aim to construct and its relationship with the linguistic choices that will potentially help them attain that effect. Meanwhile, it is also essential to raise doctoral writers' critical language awareness so that they can agentively choose from the socially available discoursal repertoire in order for them to achieve rhetorical effects and address readers' needs and expectations, and to present their scholarly identity in writing. To achieve this, it is probably necessary to include more specific information about language use in doctoral writers' academic writing course.

Another prominent finding in this dissertation study that could provide insights for doctoral academic writing is that identity is variously constructed in writing of different genres (e.g., Ivanič, 1998) written by the same writer, as demonstrated in Nancy's case. Doctoral students may have to write for various purposes during their studies, and they should be aware that writing and the identities constructed out of those texts could differ under specific situational context. Doctoral students hence can accommodate their linguistic choices to specific writing context or tasks, so that they can successfully construct identities expected by the target discourse community/ies.

In the current era of post-traditional society, individuals have more opportunities to negotiate their relationships with people from or outside their academic field (Flowerdew & Wang, 2015). It is crucial for doctoral students to engage themselves in the discourse community/ies and participate in various academic activities. Nystrand (1990, p. 5) points out that discourse communities "*inform* its speaker's discourse" (emphasis in original). The exposure they get from the activities can potentially inform their identity construction in their papers. For example, Momo's *autobiographical self* as an anonymous reviewer provided her with valuable opportunities to reflect on her own papers from a journal reviewer's perspective, which greatly helped her to adopt writing techniques preferred by the discourse community members and avoid styles that may damage the outcome of the submission. While the opportunity to engage in manuscript reviewing may not be available for a majority of doctoral students, they can participate in academic events regularly to accumulate experience as academic writers. As Nancy's and Momo's academic experience was relatively richer, they received feedback and comments from journal reviewers and conference abstract reviewers, which proved beneficial for their construction of writer identity. Further, the two cases demonstrated that guidance from senior researchers plays a critical role in helping doctoral writers construct a scholarly identity in their papers. Therefore, doctoral students can try to be exposed to more diverse academic contexts by branching out in order to receive feedback from experts through various ways. That doctoral writers' agentively seeking feedback practices (Sung, 2022) could improve their writing and identity construction. For example, doctoral writers can submit articles to journals from which they could receive feedback from anonymous reviewers, and they can attend conferences to communicate with other interested scholars who may also offer some useful feedback so that they can benefit from these activities and facilitate their identity construction as competent writers.

Individual writer's agency has been documented in the current academia to play a role in the L2 doctoral writers' socialization process. L2 doctoral students can critically learn from the practices commonly accepted in their discourse communities and gradually develop their own ways of constructing identities maturely. Advisors play an important role in empowering L2 doctoral students to exercise agency to gradually enter the core of their discourse community/ies (Li & Han, 2023). Doctoral students can communicate with advisors regularly on their academic writing to facilitate their identity construction as scholarly writers.

This dissertation study may have some implications for EAP classroom instruction. It is essential for EAP students to learn to project their identities through academic writing. To this end, EAP instructors can foster students' audience awareness, since some students may not realize the importance of readers in influencing their linguistic choices that reflect their identities through writing. Instructors can encourage students to construct identities taking into consideration the audience and communicative contexts by including the knowledge of writer identity and metadiscourse as a component in classroom instruction. With the meta-knowledge of writer identity, students can consciously construct desirable identities in their course assignments, which can lay a foundation for identity construction in other disciplinary genres beyond the EAP classroom.

5.5 Limitations of the study

As a multi-case qualitative study, this dissertation is not without limitations. First, as writer identities are constructed in situational context, which can vary depending on the purposes and goals for writing, the analytical framework used in this study, that is, metadiscourse, could only capture the identities projected in specific written genres. It is crucial for academic writers to take on multiple identities and to project identities that are desirable to community members (e.g., Kessler, 2022). Identity is always constructed, and the multifarious identities in other situational contexts, however, may be beyond the scope of this dissertation approaching writer identity with metadiscourse model.

Second, the participants came from the same linguistic and cultural background, which may not be able to reflect identity construction of L2 doctoral writers from other background. As an important strand constituting one's *autobiographical self*, doctoral writers coming from other cultural background than China could possibly provide additional insights with regard to how *autobiographical self* shape identity construction in written discourse. Further, this study did not address the diachronic developmental pattern of identity construction within a discipline due to the lack of data from a senior PhD student coming from Adam's field of study. If the data were available, it could be insightful to see how identities constructed similarly or differently by fresh PhD students and senior PhD students from the same discipline. Fourth, Adam's genres were different from the other two participants. As the other two participants' papers consisted of journal papers, if data of the same genre were available from Adam, comparisons could be made regarding identity construction in the same genres across disciplines.

Last but not least, this study did collect data concerning readers' (including journal article reviewers, advisors and course instructors) perceptions on the three doctoral students' identity construction in their papers. What was available was the readers' feedback given on the paper. As demonstrated by Zhao and Wu (2022), readers' perceptions of writers' identity may also depend on multiple factors such as their interpretations of language elements and idiosyncratic understandings of evaluative criteria. Therefore, interviews with readers would have been helpful to provide a deeper and comprehensive understanding of identities constructed in the doctoral students' papers from readers' perspective.

It also needs to be pointed out that as a qualitative multi-case study, this dissertation only reported identities constructed by three cases in their writing of various genres; hence the findings may not be able to be generalized to other populations. Not aiming to attain results generalizability, qualitative studies "produce thorough descriptions as well as a deep, rich, and contextualized understanding of human experience" (Carminati, 2018, p. 2096). However, approaching writer identity from a social constructivist perspective, which views that identity is socially constructed and it is mediated by individual's agency, the insights generated from the study could display analytical generalizability (e.g., Polit & Beck, 2010; Smith, 2018). That is, a set of findings that were generated from this dissertation study were correspondent to the established conceptualization of writer identity from the social constructivist perspective.

5.6 Suggestions for future research

Writer identity research has witnessed some development in recent years; however, we still need more research to have a deepened understanding of this construct. Since identity is not constructed in a vacuum, and contextual factors play a prominent role in the construction of identities, future study can consider how multiple contextual factors, such as culture, disciplines, writers' stages of study, audience awareness, and goals, could affect identity construction in various ways. Taking into account these factors could provide us understandings of how identity construction is informed by context in multifaceted ways. Further, identity develops diachronically, which could possibly undergo the process of non-linear and cyclical development (Ivanič, 1998); hence future research could focus on how writer identities develop over time employing a diachronic research design. Specifically, researchers can track doctoral writers' identity construction at different stages of study in their PhD journey through investigating written product their completed. This way, it will be insightful to disclose the linguistic patterns that reflect doctoral writers' identity over time, and how at each stage, contextual factors that could contribute to their identity construction. Further, according to Ivanič, writer identity constructed from texts can contribute to the transition and transformation of the socially available repertoire of writer identity; that is, *discoursal self* and *self as author* can shape *possibilities for selfhood*. Future studies can focus on how the dynamic interactions between these branches of Ivanič's conceptualization of writer identity and how *possibilities for selfhood* can be shaped by individuals' writer identity construction.

5.7 Conclusion

In this dissertation, I have presented a multi-case study including three participants who were L2 doctoral students originally coming from China and who were from two disciplines. Specifically, I examined writer identities constructed in their papers of different kinds (i.e., papers of different genres and/or papers completed at different times of their PhD studies). Metadiscourse model (Hyland, 2018) and interactional model (2005) were combined together to analyze the identities constructed in the students' papers. Further, background interviews were conducted to create profiles of the participants and interviews about the papers were conducted to uncover the variables that influenced the students' writer identities in their papers.

To answer the first research question, which examined the metadiscourse markers the doctoral writers used to construct their identities and the patterns of metadiscourse use in identity construction in their papers, I first presented the distributions of each type of the metadiscourse markers, and the results showed that the three broad categories of metadiscourse (interactive metadiscourse, stance and engagement markers) occurred in a similar pattern in all the doctoral students' papers. This demonstrated that the participants in this study, regardless of their disciplines, presented identities of writers that used explicit linguistic elements such as *endophoric markers, frame markers,* and *code gloss* to organize texts coherently, which could facilitate readers' processing of the content. The relatively less usage of stance and engagement markers denoted that it was a less conspicuous need to construct identities as academic writers positioning their stance and engaging readers into the argumentation.

The second research question concerned how the contextual variables (*possibilities for selfhood* and *autobiographical self*) and individual agency played a role in the construction of identities in the doctoral students' papers. The results showed that *possibilities for selfhood* were prominent in the shaping of writer identities in the three cases. For example, field writing conventions were found to influence the students' language choices and identity construction among all three doctoral writers. The writers' agency mattered to a lesser extent compared to *possibilities for selfhood*. However, it was found that the two doctoral writers from the hard disciplines in this study seemed to not realize the functions of the language choices they made, and their language selections were mainly based on intuition and/or perfunctory purposes, without the awareness that their identities could be projected from the language they chose.

This study has demonstrated the context-dependent nature of writer identity, which could be influenced by such variables as genres, and simultaneously individual agency played a role in the construction of writer identity. The fluid and multifaceted nature of identity was also emergent in this dissertation study. Previous studies have investigated identity construction in various groups of writers, including master's students (Supasiraprapa & De Costa, 2017), manuscript writers (Matsuda & Tardy, 2007; Tardy & Matsuda, 2009), postgraduate dissertations (Hyland, 2018), undergraduate students (Hyland, 2002a; Li & Deng, 2019), and L2 doctoral students in Hong Kong (Deng, 2012). This study focused on L2 doctoral writers in the ESL setting from two sharply distinctive disciplines, and revealed the similarities and differences in regard to metadiscourse usage that contributed to identity construction in these participants' papers. Further, issues related to L2 doctoral students' identity construction were emergent from the current study such as identity may be a foreign notion among L2 doctoral students studying in the ESL setting. Some participants may not be able to manipulate available resources to better serve the rhetorical needs of their papers, and to construct intended identities. This is worth researchers' and advisors' attention. Given this, this dissertation made some suggestions for advisors and doctoral students to consider to contribute to their identity construction as competent scholar writers. Future studies could address doctoral writers’ identity construction and provide additional insights to the understanding of this construct.

REFERENCES

Abasi, A. R. (2012). The pedagogical value of intercultural rhetoric: A report from a Persian-as-a-foreign-language classroom. *Journal of Second Language Writing*, *21*(3), 195-220.

Abasi, A. R., Akbari, N., & Graves, B. (2006). Discourse appropriation, construction of identities, and the complex issue of plagiarism: ESL students writing in graduate school. *Journal of Second Language Writing, 15*, 102-117.

Adams, G. R., & Marshall, S. K. (1996). A developmental social psychology of identity: Understanding the person-in-context. *Journal of adolescence*, *19*(5), 429-442.

Ahearn, L. M. (2001). Language and agency. *Annual Review of Anthropology, 30,* 109-137.

Ahern, K., & Le Brocque, R. (2005). Methodological issues in the effects of attrition: Simple solutions for social scientists. *Field methods*, *17*(1), 53-69.

Aitchison, C., Catterall, J., Ross, P., & Burgin, S. (2012). 'Tough love and tears': Learning doctoral writing in the sciences. *Higher Education Research & Development*, *31*(4), 435-447.

Alkhanbooli, B. R. (2018). *Investigating L2 Writers' Investment Toward Constructing a Successful Writer Identity: Case Studies of Arab Students from Saudi Arabia Studying in US* (Doctoral dissertation, University of Memphis).

Amineh, R. J., & Asl, H. D. (2015). Review of constructivism and social constructivism. *Journal of Social Sciences, Literature and Languages*, *1*(1), 9-16.

Anderson, T. (2017). The doctoral gaze: Foreign PhD students' internal and external academic discourse socialization. *Linguistics and Education*, *37*, 1-10.

Angélil-Carter, S. (1995). *Uncovering plagiarism in academic writing: Developing authorial voice within multivoiced text* (master's dissertation, Rhodes University).

Angélil-Carter, S. (1997). Second language acquisition of spoken and written English: Acquiring the skeptron. *TESOL quarterly*, *31*(2), 263-287.

Atkinson, D. (2003). L2 writing in the post-process era: Introduction. *Journal of Second Language Writing*, *12*(1), 3–15.

Asprillia, A., & Hardjanto, T. D. (2020). WHO AM "I" IN ACADEMIC WRITING?: THE STUDY OF AUTHORIAL IDENTITY. *LLT Journal: A Journal on Language and Language Teaching*, *23*(1), 131-143.

Bakhtin, M. M. (1986). The Problem of Speech Genres. In C. Emerson and M. Holquist (Eds.), *Speech Genres and other Late Essays* (pp. 60-102). Austin: University of Texas Press.

Bartholomae, D. (1986). Inventing the university. *Journal of Basic Writing, 5*(1), 4-23.

Benwell, B., & Stokoe, E. (2006). *Discourse and identity*. Edinburgh University Press.

Berger, P. L. (1963). Invitation to sociology. *A Humanistic Perspective.* New York.

Berger, P. L., & Luckmann, T. (1966). *The social construction of reality.* Penguin Books.

Bernard, H., & Ryan, G. (2010). *Analyzing Qualitative Data: Systematic Approaches.* Thousand Oaks CA, London: SAGE Publications.

Bhatia, V. K. (2004). *Worlds of written discourse: A genre-based view*. A&C Black.

Biglan, A. (1973). The characteristics of subject matter in different academic areas. *Journal of applied psychology, 57*(3), 195-203.

Block, D. (2007). The rise of identity in SLA research, post Firth and Wagner (1997). *The Modern language journal*, *91*, 863-876.

Bogdan, R. C., & Biklen, S. K. (1992). *Qualitative research for education: An introduction to theory and methods*. Boston: Allyn & Bacon.

Bosma, H. A., Graafsma, T. L. G., Grotevant, H. D. & de Levita, D. J. (1994) (Eds). *Identity and development: an interdisciplinary approach.* Newbury Park, CA: Sage.

Botelho de Magalhães, M., Cotterall, S., & Mideros, D. (2019). Identity, voice and agency in two EAL doctoral writing contexts. *Journal of Second Language Writing*, *43*, 4-14.

Bourdieu, P. (1986). The forms of capital. In John G. Richardson (Ed), *Hand-book of theory and research for the sociology of education* (pp. 241–258). New York: Greenwood Press.

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Thousand Oaks, CA: Sage.

Boz, C. (2006). *Establishing an academic identity: Second language writers and the institution* (Doctoral dissertation, Sheffield Hallam University).

Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse studies*, *7*(4-5), 585-614.

Burgess, A. (2012). “I don’t want to become a China Buff”: Temporal dimensions of the discoursal construction of writer identity. *Linguistics and Education*, *23*(3), 223-234.

Burgess, A., & Ivanič, R. (2010). Writing and being written: Issues of identity across timescales. *Written Communication*, *27*(2), 228-255.

Burke, S. B. (2010). *The construction of writer identity in the academic writing of Korean ESL students: A qualitative study of six Korean students in the US*. Doctoral dissertation. Indiana University of Pennsylvania.

Butler, J. (1990), *Gender Trouble: Feminism and the Subversion of Identity*, New York: Routledge.

Butler, J. (1997). *The psychic life of power: Theories in subjection.* Stanford: Stanford University Press.

Cameron, D. (2001). *Working with spoken discourse*. London: Sage.

Canagarajah, S. (2000). Negotiating ideologies through English: Strategies from the periphery. In T. Ricento, (Ed.), *Ideology, politics and language policies: Focus on English* (pp. 121–132). Philadelphia: John Benjamins.

Canagarajah, S. (2004). Multilingual writers and the struggle for voice in academic discourse. In A. Pavlenko & A. Blackledge (Eds.), *Negotiation of identities in multilingual contexts*, (pp. 266-289). Clevedon, UK: Multilingual Matters.

Canagarajah, S., & Matsumodo, Y. (2016). Negotiating Voices in Translingual Literacies: From literacy regimes to contact zones. *Journal of Multilingual and Multicultural Development, 38*(5), 1–17.

Candlin, C. N., & Sarangi, S. (2003). Introduction: Trading between reflexivity and relevance: New challenges for applied linguistics. *Applied Linguistics*, *24*(3), 271-285.

Carminati, L. (2018). Generalizability in qualitative research: A tale of two traditions. *Qualitative health research*, *28*(13), 2094-2101.

Casanave, C. P. (2019). Performing expertise in doctoral dissertations: Thoughts on a fundamental dilemma facing doctoral students and their supervisors. *Journal of Second Language Writing*, *43*, 57-62.

Cassell, J. (1978). Risk and benefit to subjects of fieldwork. *The American Sociologist*, 134-143.

Castelló, M., Iñesta, A., & Corcelles, M. (2013). Learning to write a research article: Ph. D. students' transitions toward disciplinary writing regulation. *Research in the Teaching of English*, 442-477.

Castillo-Hajan, B., Hajan, B. H., & Marasigan, A. C. (2019). Construction of Second Language Writer Identity in Student Persuasive Essays: A Metadiscourse Analysis. *Online Submission*, *21*, 36-60.

Chang, P., & Schleppegrell, M. (2016). Explicit learning of authorial stance-taking by L2 doctoral students. *Journal of Writing Research*, *8*(1), 49-80.

Charles, M. (2006). Revealing and obscuring the writer's identity: Evidence from a corpus of theses. In R. Kiely, G. Clibbon, P. Rea-Dickins and H. Woodfield (Eds.), *Language, culture and identity in applied linguistics* (pp. 147-161). London: Equinox Publishing Ltd.

Cherry, R. D. (1998). Ethos versus persona: Self-representation in written discourse. *Written communication, 15*(3), 384-410.

Chiu, Y. L. T. (2016). 'Singing your tune': Genre structure and writer identity in personal statements for doctoral applications. *Journal of English for Academic Purposes*, *21*, 48-59.

Clark, R., & Ivanič, R. (1997). *The politics of writing*. Routledge.

Clarke, M. (2009). The ethico-politics of teacher identity. *Educational Philosophy and Theory, 41*(2), 185-200.

College Factual. (2022, November 8). *University of South Florida - Main Campus International*. Retrieved November 12, 2022. https://www.collegefactual.com/colleges/university-of-south-florida-main-campus/student-life/international/#secOverallPopularity

Crompton, P. (1997). Hedging in academic writing: Some theoretical problems. *English for specific purposes*, *16*(4), 271-287.

Cumming, A., Busch, M., & Zhou, A. (2002). Investigating learners' goals in context of adult second language writing. In S. Ransdell & M. Barbier (Eds.), *New directions for research in L2 writing* (pp.189-208). Dordrecht, Netherlands: Kluwer Academic Publisher.

Danis, N. (2022). A corpus analysis of disciplinary identity in evaluative journal articles: A Systemic Functional Linguistics approach. *English for Specific Purposes*, *68*, 87-101.

De Costa, P. I., Randez, R. A., Her, L., & Green-Eneix, C. A. (2021). Navigating ethical challenges in second language narrative inquiry research. *System*, *102*, 102599.

de Fina, A. E., Schiffrin, D. E., & Bamberg, M. E. (2006). *Discourse and identity*. Cambridge University Press.

Deng, L. (2012). Academic identity construction in writing the discussion & conclusion section of L2 theses: Case studies of Chinese social science doctoral students. *Chinese Journal of Applied Linguistics*, *35*(3), 301-323.

Denzin, Y. S. & Lincoln, N. K (2005). *The Sage handbook of qualitative research*. Sage.

Dornyei, Z. (2007) *Research Methods in Applied Linguistics*. Oxford University Press.

du Gay, P., Evans, J., & Redman, P. (2000). *Identity: A reader.* London: Sage.

Duff, P. (2008). *Case study research in applied linguistics*. Routledge.

Duff, P. A., & Uchida, Y. (1997). The negotiation of teachers' sociocultural identities and practices in postsecondary EFL classrooms. *Tesol Quarterly*, *31*(3), 451-486.

Eisner, E. W. (1991). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice.* New York: Macmillan.

Elias, N. (1998). *On civilization, power, and knowledge: Selected writings* (Vol. 1). University of Chicago Press.

Ely, M. (with Anzul, M., Friedman, T., Garner, D., & Steinmetz, A. M.). (1991). *Doing qualitative research: Circles within circles.* London: Falmer Press.

Fairclough, N. (1989). *Language and Power.* London: Longma.

Fairclough, N. (1992). Intertextuality in critical discourse analysis. *Linguistics and Education*, *4*(3–4), 269–293.

Fairclough, N. (2003). *Analysing discourse: Textual analysis for social research*. Psychology Press.

Fazel, I., & Shi, L. (2015). Citation behaviors of graduate students in grant proposal writing. *Journal of English for Academic Purposes*, *20*, 203-214.

Flick, U. (2018). Triangulation in data collection. In Flick, U. (Ed). *The SAGE handbook of qualitative data collection* (pp. 527-544). SAGE.

Flowerdew, J. (2001). Attitudes of journal editors to nonnative speaker contributions. *TESOL quarterly*, *35*(1), 121-150.

Flowerdew, J., & Wang, S. H. (2015). Identity in academic discourse. *Annual Review of Applied Linguistics*, *35*, 81-99.

Frick, B. L., & Brodin, E. M. (2020). A return to Wonderland: Exploring the links between academic identity development and creativity during doctoral education. *Innovations in Education and Teaching International*, *57*(2), 209-219.

Gall, J., Gall, M. D., & Borg, W. T. (2005). *Applying educational research* (5th ed.). Boston: Pearson Education.

Gee, J. P. (2000). Chapter 3: Identity as an analytic lens for research in education. *Review of research in education*, *25*(1), 99-125.

Gee, J. (2015). *Social linguistics and literacies: Ideology in discourses*. Routledge.

Gee, J. P. (2017). Identity and diversity in today's world. *Multicultural Education Review*, *9*(2), 83-92.

Giddens, A. (1991). *Modernity and self-identity: Self and society in the late modern age*. Redwood City, CA: Stanford University Press.

Goetz, J. P., & LeCompte, M. D. (1984). *Ethnography and Qualitative Design in Educational Research.* Orlando, Fla.: Academic Press.

Green, B. (2005). Unfinished business: Subjectivity and supervision. *Higher Education Research and Development 24*: 151-163.

Griffin, L. (2006). Who or what is the students' audience? The discoursal construction of audience identity in undergraduate assignments. In R. Kiely., G. Clibbon., P. Rea-Dickins,. & H. Woodfield. (Eds.), *Language, culture and identity in applied linguistics* (pp. 133-146). London: Equinox Publishing Ltd.

Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. Jossey-Bass.

Hall, J. K. (1995). (Re)creating our worlds with words: A sociohistorical perspective of face-to-face interaction. *Applied Linguistics, 16,* 206–232.

Halliday, M. A. K. (1985). *Introduction to functional grammar.* London: Arnold (2nd ed., 1994).

Halliday, M., & Matthiessen, C. (2014). *An introduction to functional grammar (3rd edition)*. Routledge.

Halliday, M. A. K., & Martin, J. R. (2003). *Writing science: Literacy and discursive power*. Routledge.

Hammersley, M., & Atkinson, P. (1983). *Ethnography: Principles in Practice.* London: Routledge.

He, A. W. (1995). Co-constructing institutional identities: The case of student counselees. *Research on Language and Social Interaction, 28,* 213–231.

Hirtle, J. S. P. (1996). Social constructivism. *English Journal*, *85*(1), 91-92.

Hirvela, A., & Belcher, D. (2001). Coming back to voice: The multiple voices and identities of mature multilingual writers. *Journal of second language writing*, *10*(1-2), 83-106.

Howard, R. M. (1995). Plagiarisms, authorship, and the academic death penalty. *College English, 57*(7), 788–806.

Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied linguistics*, *20*(3), 341-367.

Hyland, K. (2000). *Disciplinary Discourses: Social Interactions in Academic Writing.* London: Longman.

Hyland, K. (2001). Bringing in the reader: Addressee features in academic articles. *Written communication*, *18*(4), 549-574.

Hyland, K. (2002a). Authority and invisibility: Authorial identity in academic writing. *Journal of pragmatics*, *34*(8), 1091-1112.

Hyland, K. (2002b). Directives: Argument and engagement in academic writing. *Applied linguistics*, *23*(2), 215-239.

Hyland, K. (2002c). What do they mean? Questions in academic writing. *Text & Talk*, *22*(4), 529-557.

Hyland, K. (2004a). *Disciplinary discourses, Michigan classics ed.: Social interactions in academic writing*. University of Michigan Press.

Hyland, K. (2004b). Disciplinary interactions: Metadiscourse in L2 Postgraduate writing. *Journal of Second Language Writing, 13*, 133-151.

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse studies*, *7*(2), 173-192.

Hyland, K. (2007). Applying a gloss: Exemplifying and reformulating in academic discourse. *Applied linguistics*, *28*(2), 266-285.

Hyland, K. (2010). Community and individuality: Performing identity in applied linguistics. *Written communication*, *27*(2), 159-188.

Hyland, K. (2012a). *Disciplinary Identities: Individuality and Community in Academic Discourse.* London: Oxford University Press.

Hyland, K. (2012b). Undergraduate understandings: Stance and voice in final year reports. In K. Hyland, & C. S. Guinda (Eds.). *Stance and voice in written academic genres* London: Palgrave Macmillan pp. 134–150.

Hyland, K. (2014). Disciplinary discourses: Writer stance in research articles. In C. N. Candlin & K. Hyland (Eds.). *Writing: Texts, processes and practices* (pp. 99-121). Routledge.

Hyland, K. (2015). Genre, discipline and identity. *Journal of English for academic purposes*, *19*, 32-43.

Hyland, K. (2016). Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing, 31,* 58–69.

Hyland, K. (2017). Metadiscourse: What is it and where is it going?. *Journal of pragmatics*, *113*, 16-29.

Hyland, K. (2018). *Metadiscourse: Exploring interaction in writing*. Bloomsbury Publishing.

Hyland, K., & Jiang, F. K. (2016). "We must conclude that…": A diachronic study of academic engagement. *Journal of English for Academic Purposes*, *24*, 29-42.

Hyland, K., & Jiang, F. K. (2018). "In this paper we suggest": Changing patterns of disciplinary metadiscourse. *English for Specific Purposes*, *51*, 18-30.

Hyland, K., & Jiang, F. K. (2022). Metadiscourse choices in EAP: an intra-journal study of JEAP. *Journal of English for Academic Purposes*, *60*, 101165.

Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied linguistics*, *25*(2), 156-177.

Hyland, K., & Zou, H. J. (2020). In the frame: Signalling structure in academic articles and blogs. *Journal of Pragmatics*, *165*, 31-44.

Inouye, K., & McAlpine, L. (2019). Developing academic identity: A review of the literature on doctoral writing and feedback. *International Journal of Doctoral Studies*, *14*, 1-31.

Işık-Taş, E. E. (2018). Authorial identity in Turkish language and English language research articles in Sociology: The role of publication context in academic writers' discourse choices. *English for Specific Purposes*, *49*, 26-38.

Ivanič, R. (1994). I is for interpersonal: Discoursal construction of writer identities and the teaching of writing. *Linguistics and education*, *6*(1), 3-15.

Ivanič, R. (1998). *Writing and identity: The discoursal construction of identity in academic writing*. John Benjamins.

Ivanič, R. (2004). Discourses of writing and learning to write. *Language and Education, 18*, 220-245.

Ivanič, R. (2006). Language, learning and identification. In R. Kiely; G. Clibbon; P. Rea-Dickins and H. Woodfield (Eds.), *Language, culture and identity in applied linguistics* (pp. 7-29). London: Equinox Publishing Ltd.

Ivanič, R., & Camps, D. (2001). I am how I sound: Voice as self-representation in L2 writing. *Journal of second language writing*, *10*(1-2), 3-33.

Ivanič, R., & Weldon, S. (2014). Researching the writer-reader relationship. In C. N. Candlin and K. Hyland (Eds.). *Writing: Texts, processes and practices* (pp. 168-192). Routledge.

Jenkins, R. (2008). *Social Identity* (3rd ed.). London: Routlege.

Johnson, D. M. (1992). *Approaches to Research in Second Language Learning*. New York: Longman Publishing Group.

Kamler, B. (2008). Rethinking doctoral publication practices: Writing from and beyond the thesis. *Studies in higher education*, *33*(3), 283-294.

Kamler, B., & Thomson, P. (2006). *Helping doctoral students write: Pedagogies for supervision*. Oxon: Routledge.

Kamler, B., & Thomson, P. (2014). *Helping doctoral students write: Pedagogies for supervision*. Oxon: Routledge.

Kessler, M. (2022). Prospective English language teachers' projected identities in personal statements plus experienced teachers' reactions. *International Journal of Applied Linguistics*, *32*(1), 76-93.

Kim, B. (2001). Social Constructivism. In M. Orey (Ed.), *Emerging perspectives on learning, teaching, and technology* (pp. 55-61). North Charleston: CreateSpace.

Kim, T. Y., Baba, K., & Cumming, A. (2006). Goals, motivations, and identities of three students writing in English. In A. Cumming (Ed.), *Goals for academic writing: ESL students and their instructors* (pp. 125-141). Philadelphia: John Benjamins.

Kristeva, J. (1980). *Desire in language: a semiotic approach to literature and art.* New York: Columbia University Press.

Kuteeva, M. (2011). Wikis and academic writing: Changing the writer–reader relationship. *English for Specific Purposes*, *30*(1), 44-57.

Laband, D. N., & Tollison, R. D. (2000). Intellectual collaboration. *Journal of Political economy*, *108*(3), 632-662.

Langum, V., & Sullivan, K. P. (2017). Writing academic English as a doctoral student in Sweden: Narrative perspectives. *Journal of Second Language Writing*, *35*, 20-25.

Lee, J. J., Hitchcock, C., & Casal, J. E. (2018). Citation practices of L2 university students in first-year writing: Form, function, and stance. *Journal of English for Academic Purposes*, *33*, 1-11.

Lee, J. J., & Casal, J. E. (2014). Metadiscourse in results and discussion chapters: A cross-linguistic analysis of English and Spanish thesis writers in engineering. *System*, *46*, 39-54.

Li, X. M. (1992). *A celebration of tradition or of self? An ethnographic study of teachers' comments on student writing in America and in China.* Doctoral dissertation. University of New Hampshire.

Li, Y., & Deng, L. (2019). I am what I have written: A case study of identity construction in and through personal statement writing. *Journal of English for Academic Purposes*, *37*, 70-87.

Li, F., & Han, Y. (2023). Chinese international students' identity (re) construction mediated by teacher feedback: Through the lens of academic discourse socialisation. *Journal of English for Academic Purposes*, *61*, 101211.

Lichtman, M. (2012). *Qualitative research in education: A user's guide (3rd ed.)*. Sage publications.

Maher, D., Seaton, L., McMullen, C., Fitzgerald, T., Otsuji, E., & Lee, A. (2008). 'Becoming and being writers': The experiences of doctoral students in writing groups. *Studies in continuing education*, *30*(3), 263-275.

Markee, N. (2013). Emic and etic in qualitative research. *The encyclopedia of applied linguistics,* 1-4.

Mathison, S. (1988) 'Why triangulate?'. *Educational Researcher,* 17(2), 13–17.

Matsuda, P. K. (2001). Voice in Japanese written discourse: Implications for second language writing. *Journal of Second language writing*, *10*(1-2), 35-53.

Matsuda, P. K. (2015). Identity in written discourse. *Annual Review of Applied Linguistics*, *35*, 140-159.

Matsuda, P. K., & Tardy, C. M. (2007). Voice in academic writing: The rhetorical construction of author identity in blind manuscript review. *English for Specific Purposes*, *26*(2), 235-249.

MAXQDA Products. Available at: http://www.maxqda.com/products/maxqda. Oaks, CA: Sage.

Maxwell, J. A. (1996). *Qualitative Research Design*. London: Sage.

McKinley, J. (2015). Critical argument and writer identity: Social constructivism as a theoretical framework for EFL academic writing. *Critical inquiry in language studies*, *12*(3), 184-207.

Merriam, S. H. (1988). *Case study research in education: A qualitative approach*. San Francisco, Jossey-Bass.

Michael, M. (1996). *Constructing identities: The social, the nonhuman and change*. Sage.

Moghaddam, A. (2006). Coding issues in grounded theory. *Issues in educational research*, *16*(1), 52-66.

Morton, J., Storch, N., & Thompson, C. (2015). What our students tell us: Perceptions of three multilingual students on their academic writing in first year. *Journal of Second Language Writing*, *30*, 1-13.

Norton, B. (1997). Language, identity, and the ownership of English. *TESOL Quarterly, 31*, 409–429.

Norton, B., & Gao, Y. H. (2008). Identity, investment, and Chinese learners of English. *Journal of Asian Pacific Communication, 18*, 109–120.

Nystrand, M. (1990). Sharing words: The effects of readers on developing writers. *Written communication*, *7*(1), 3-24.

Ochs, E. (1993). Constructing social identity: A language socialization perspective. *Research on Language and Social Interaction, 26,* 287–306.

Ouellette, M. A. (2008). Weaving strands of writer identity: Self as author and the NNES "plagiarist". *Journal of Second Language Writing*, *17*(4), 255-273.

Paltridge, B. (2004). Academic writing. *Language teaching*, *37*(2), 87-105.

Paré, A. (2019). Re-writing the doctorate: New contexts, identities, and genres. *Journal of Second Language Writing*, *43*, 80-84.

Paré, A., Starke-Meyerring, D., & McAlpine, L. (2011). Knowledge and identity work in the supervision of doctoral student writing: Shaping rhetorical subjects. In Starke-Meyerring, D., Paré, A., Artemeva, N., Horne, M. and Yousoubova, L. (Eds.), *Writing in knowledge societies* (pp. 215-236). Parlor Press.

Pavlenko, A. (2000). Access to linguistic resources: Key variable in second language learning. *Sociolinguistic Studies*, 85-105.

Pavlenko, A. (2001). "In the world of the tradition, I was unimagined": Negotiation of identities in cross-cultural autobiographies. *International Journal of Bilingualism*, *5*(3), 317-344.

Pavlenko, A., & Blackledge, A. (2004). *Negotiation of identities in multilingual contexts*. Clevedon, UK: Multilingual Matters.

Polit, D. F., & Beck, C. T. (2010). Generalization in quantitative and qualitative research: Myths and strategies. *International journal of nursing studies*, *47*(11), 1451-1458.

Prior, P. (2001). Voices in text, mind, and society: Sociohistoric accounts of discourse acquisition and use. *Journal of second language writing, 10*(1-2), 55-81.

Ricento, T. (2005). Considerations of identity in L2 learning. *Handbook of research in second language teaching and learning*, *1*, 895-910.

Richards, K. (2009). Interviews. In Heigham, J. and Croker, R.A. (Eds.), *Qualitative Research in Applied Linguistics: A practical introduction.* Palgrave Macmillan.

Russell, D. (1997). Rethinking genre in school and society: An activity theory analysis. *Written Communication, 14(4)*, 504-554.

Samraj, B. (2002). Texts and contextual layers: academic writing in content courses. In A. Johns (Ed.), *Genre in the classroom: Multiple perspectives* (pp. 163-176). Mahwah, NJ: Lawrence Erlbaum Associates.

Sayer, A. (1997). Essentialism, social constructionism, and beyond. *The Sociological Review*, *45*(3), 453-487.

Schaffner, C. (1996) Editorial. In Schajjner, C. and Kelly-Holmes, H. (1996) *Discourse and Ideologies*. Multilingual Matters.

Scollon, R. (1994). As a matter of fact: The changing ideology of authorship and responsibility in discourse. *World Englishes*, *13*(1), 33-46.

Scott, M. (1999). Agency and subjectivity in student writing. In C. Jones, J. Turner and B. Street (Eds.), *Students Writing in the University: cultural and epistemological issues* (pp. 171-192). Amsterdam: Benjamins.

Sealey, A., & Carter, B. (2004). *Applied linguistics as social science*. Continuum.

Sewell, W.H. (1992). A theory of structure: Duality, agency and transformation. *American Journal of Sociology*, 98, 1–29.

Shanahan, M. C. (2009). Identity in science learning: Exploring the attention given to agency and structure in studies of identity. *Studies in Science Education*, *45*(1), 43-64.

Shen, F. (1989). The classroom and the wider culture: Identity as a key to learning English composition. *College composition and communication*, *40*(4), 459-466.

Shi, L. (2011). Common knowledge, learning, and citation practices in university writing. *Research in the Teaching of English*, 308-334.

Shi, L. (2012). Rewriting and paraphrasing source texts in second language writing. *Journal of Second Language Writing*, *21*(2), 134-148.

Silva, T. & Matsuda, P. K. (2002). Writing. In N. Schmitt (ed.), *An introduction to applied linguistics*. London: Arnold, 252–266.

Smith, B. (2018). Generalizability in qualitative research: Misunderstandings, opportunities and recommendations for the sport and exercise sciences. *Qualitative research in sport, exercise and health*, *10*(1), 137-149.

Starfield, S. (2002). "I'm a second-language English speaker": Negotiating writer identity and authority in sociology one. *Journal of language, identity, and education*, *1*(2), 121-140.

Starfield, S. (2004). “Why does this feel empowering?” In B. Norton & K. Toohey (Eds.), *Critical pedagogies and language learning* (pp. 138-157). Cambridge: Cambridge University Press.

Starfield, S., & Ravelli, L. J. (2006). “The writing of this thesis was a process that I could not explore with the positivistic detachment of the classical sociologist”: Self and structure in New Humanities research theses. *Journal of English for Academic Purposes*, *5*(3), 222-243.

Stetsenko, A., & Arievitch, I. (1997). Constructing and deconstructing the self: Comparing post-Vygotskian and discourse-based versions of social constructivism. *Mind, Culture, and Activity, 4,* 160–173.

Stock, I., & Eik-Nes, N. L. (2016). Voice features in academic texts–A review of empirical studies. *Journal of English for Academic Purposes*, *24*, 89-99.

Strauss, A. L., & Corbin, J. M. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory (2nd ed.).* Thousand Oaks: Sage Publications.

Sung, C. C. M. (2022). Agency and feedback-seeking: Academic English socialization of L2 students in Hong Kong. *Language and Education*, 1-19.

Supasiraprapa, S., & De Costa, P. I. (2017). Metadiscourse and identity construction in teaching philosophy statements: A critical case study of two MATESOL students. *TESOL Quarterly*, *51*(4), 868-896.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Swales, J. (2014). The concept of discourse community. In Wardle, E. & Downs, D (Eds). Writing about writing: A college reader (pp. 215-228). Boston: Bedford/St. Martins.

Swales, J. M., & Feak, C. B. (2004). *Academic writing of graduate students: Essential tasks and skills.* Ann Arbor: University of Michigan Press.

Taj, S. R. (2017). *Saudi Students' Social Identity and Their Identity in Academic Writing: A Qualitative Study of Saudi Students in the UK* (Doctoral dissertation, University of Leicester).

Tang, R., & John, S. (1999). The 'I' in identity: Exploring writer identity in student academic writing through the first person pronoun. *English for specific purposes*, *18*, S23-S39.

Tardy, C. (2016). 16 Voice and identity. *Handbook of second and foreign language writing*, *11*, 349-364.

Tardy, C. M., & Matsuda, P. K. (2009). The construction of author voice by editorial board members. *Written Communication, 26*(1), 32-52.

Tessuto, G. (2008). Writer identity in the Introduction section of academic law research articles: Exploring metadiscourse strategies. *Linguistica e Filologia, 27,* 39-58.

Thompson, P. (2012). Achieving a voice of authority in PhD theses. In K. Hyland, & C. S. Guinda (Eds.). *Stance and voice in written academic genres* (pp. 119–133). UK: Palgrave MacMillan.

Thorton, M. A. (2003). *Writer identity as a discourse theory*. [Master's thesis, Tarleton State University]. ProQuest Dissertations Publishing.

Tracy, K., & Robles, J. S. (2013). *Everyday talk: Building and reflecting identities*. Guilford Press.

Vygotsky, L. S. (1978). *Mind in society.* Cambridge: Harvard University Press.

Wang, M., & Parr, G. (2021). A Chinese doctoral student's experience of L2 English academic writing in Australia: Negotiating practices and identities. *Journal of English for Academic Purposes*, *49*, 100944.

Webber, P. (1994). The function of questions in different medical English genres. *English for Specific Purposes* 13: 257–68.

Weidman, J. C., & Stein, E. L. (2003). Socialization of doctoral students to academic norms. *Research in higher education*, *44*, 641-656.

Wenger, E. (1999). *Communities of practice: Learning, meaning, and identity*. Cambridge university press.

Williams, R. (2000). *Making identity matter: identity, society and social interaction*. Routledge Cavendish.

Worden, D. (2018). Balancing stability and flexibility in genre-based writing instruction: A case study of a novice L2 writing teacher. *Journal of Second Language Writing*, *42*, 44-57.

Wu, D., & Buripakdi, A. (2021). Writer Identity Construction in EFL Doctoral Thesis Writing. *GEMA Online Journal of Language Studies*, *21*(3), 16-36.

Xu, L. (2017). Written feedback in intercultural doctoral supervision: A case study. *Teaching in Higher Education*, *22*(2), 239-255.

Xu, L., & Zhang, L. J. (2019). L2 doctoral students' experiences in thesis writing in an English-medium university in New Zealand. *Journal of English for Academic Purposes*, *41*, 100779.

Yin, R. K. (2018). *Case study research and applications: Design and methods*. Sage publications.

Zhao, C. G. (2012). Authorial voice in second language writing. *The Encyclopedia of Applied Linguistics*, 1-6.

Zhao, C. G. (2017). Voice in timed L2 argumentative essay writing. *Assessing writing*, *31*, 73-83.

Zhao, C. G., & Llosa, L. (2008). Voice in high-stakes L1 academic writing assessment: Implications for L2 writing instruction. *Assessing Writing*, *13*(3), 153-170.

Zhao, C. G., & Wu, J. (2022). Perceptions of authorial voice: Why discrepancies exist. *Assessing Writing*, *53*, 100632.

Zhang, Y. O., & Hyland, K. (2021). Elements of doctoral apprenticeship: Community feedback and the acquisition of writing expertise. *Journal of Second Language Writing*, *53*, 100835.

Zhu, W. (2004). Faculty views on the importance of writing, the nature of academic writing, and teaching and responding to writing in the disciplines. *Journal of second language Writing*, *13*(1), 29-48.

Zou, H., & Hyland, K. (2019). Reworking research: Interactions in academic articles and blogs. *Discourse studies*, *21*(6), 713-733.

Zou, H. J., & Hyland, K. (2020). "Think about how fascinating this is": Engagement in academic blogs across disciplines. *Journal of English for Academic Purposes*, *43*, 100809.

Zou, H., & Hyland, K. (2022). Stance in academic blogs and three-minute theses. *International Journal of Applied Linguistics*.

APPENDIX A: BACKGROUND INTERVIEW GUIDELINES

Personal background information

1. Please tell me your background information, including age, doctoral program, and department.
2. Why did you choose to pursue a PhD degree in the US? How do you feel about the academic environment in the US? Compared to your home country or other country/ies you have studied in, what are the similarities and differences?
3. What was your undergrad major? What was your specialization in your master program?
4. What is your plan after graduation?
5. Did you take TOEFL or IELTS? What was your score? And score for the writing section?
6. In both your undergrad and master's program, what courses related to academic writing did you take?

Academic writing practice and identity construction

7. How often do you usually write in English? What kinds of writings are they? For what purposes do you write?
8. Please tell me about your English learning and writing experience. How did that influence your academic writing in English?
9. What are your usual steps of writing in English? What do you do first, second, and so on?
10. Do you use translation when writing your papers?
11. What aspects of writing did you find easy or difficult?

12. What specific strategies do you use in your writing? What do you do when you have trouble writing?
13. Have you ever written academic papers in Chinese? What are the differences between writing in Chinese vs. English?
14. What part/s of English writing do you think is difficult? What is the reason for that? Is it difficult for you to write for academic purposes in English?
15. Do you consider yourself to be a successful writer in English? Why or why not? How would you describe yourself as an English writer?

APPENDIX B: INTERVIEW ON PAPERS GUIDELINES

1. When you wrote the paper, what was your purpose?
2. Did you have any support from people that helped with your writing assignments? (e.g., friends, professors, or tutors)?
3. Have you thought about your reader/s when writing the paper? Who were they? How did they influence your writing?
4. What were your instructor/advisor/blind reviewer's opinions about your writing? What about your peers' opinions? Other people's? How do you feel about those opinions?
5. What kind/s of impressions did your paper bring to your readers? After rereading your own writing, what kinds of impressions do you have? Does it reflect who you are (your personal, social, or ethnic identity)?
6. What were your considerations when making an argument or when you disagreed with an opinion?
7. Are there similarities and differences that you noticed that existed in your papers you wrote at different times? What are they? What caused the similarities and differences?
8. What kinds of effect Chinese writing brings to your English writing, if any?
9. What other factors have you thought about during writing that influenced your choice of certain expressions?
10. What aspects does your teachers' feedback usually focus on? Do you think that helps your writing? On what aspects of writing do you wish your teachers to give feedback?

11. Here, when you used words such as “definitely, significantly”, how confident or certain were you?
12. Here, why did you use expressions such as “may, might, possible”?
13. I noticed at the end of the introduction section, you would have a paragraph that was about the overview of the paper. Is that something people all follow in the discipline?
14. Here when you used adjectives such as “surprising, interesting”, what were the considerations for that?
15. Please tell me why you chose this specific word and/or sentence structure in your essay. Did you realize any differences in your different papers?
16. Please tell me why you cited this source in this way in your paper (if any). Did you realize any differences in your different papers?

APPENDIX C: THE INSTITUTIONAL REVIEW BOARD (IRB) EXEMPTION LETTER

EXEMPT DETERMINATION

February 15, 2021

Beibei Ren
[redacted]

Dear Ms. Beibei Ren:

On 2/14/2021, the IRB reviewed and approved the following protocol:

Application Type:	Initial Study
IRB ID:	STUDY002159
Review Type:	Exempt 2, 4
Title:	L2 writer identity construction in academic written discourse: A multi-case study
Protocol:	• Protocol ;

The IRB determined that this protocol meets the criteria for exemption from IRB review.

In conducting this protocol, you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Please note, as per USF policy, once the exempt determination is made, the application is closed in BullsIRB. This does not limit your ability to conduct the research. Any proposed or anticipated change to the study design that was previously declared exempt from IRB oversight must be submitted to the IRB as a new study prior to initiation of the change. However, administrative changes, including changes in research personnel, do not warrant a modification or new application.

Ongoing IRB review and approval by this organization is not required. This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these activities impact the exempt determination, please submit a new request to the IRB for a determination.

Sincerely,

Jennifer Walker
IRB Research Compliance Administrator

Institutional Review Boards / Research Integrity & Compliance
FWA No. 00001669
University of South Florida / 3702 Spectrum Blvd., Suite 165 / Tampa, FL 33612 / 813-974-5638

Page 1 of 1