Interpreting students' ability in self-assessment.

Eunjin Heo

Thesis submitted in partial fulfilment of the requirements

for the degree of Master of Research

Department of Linguistics

Macquarie University

July 2018

Contents

Abstract

Interpreting students need to be able to self-assess their performance during and after their education to effectively develop their interpreting skills for professional practice. Although there is a growing interest in self-assessment in interpreter education research, it is as yet an under-researched area, especially with regard to the interpreting students' ability to self-assess.

This thesis reports the results of a case study that examined interpreting students in a graduate program in Australia and its joint offshore program in Korea. Using a mixed methods approach, the students' reflective journals, interviews with teachers, the students' self-rated scores and teachers' grades were examined with the aim of understanding students' ability to assess their performance and the relationship between their self-assessment ability and interpreting performance gains over the course of their study. In addition, the impact of teacher instruction on self-assessment was examined.

The findings of this research indicate that interpreting students' self-ratings were different from their teacher-rated scores, with the majority of students underestimating their performance. In addition, their self-assessment was largely focused on the negative aspects of their performance and it appeared they might not be able to apply some of the scoring criteria adequately in making judgements about their performance. It is also apparent that teachers make assumptions about their students' ability to apply rating criteria for self-assessment.

Overall, the findings suggest that self-assessment is a complex task, that students cannot intuitively develop skills to self-assess while they are acquiring interpreting skills, and that their ability depends to some extent on the guidance they receive from their teachers.

Keywords: interpreter education, self-assessment, criteria, interpreting students' ability to self-assess.

Statement of Candidate

This work has not previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Ethics approval was granted by the Macquarie University Faculty of Human Sciences Human Research Ethics Sub-Committee on 30th October 2017, reference number: 5201700979.

Eunjin Heo
6 July 2018

List of Figures

List of Tables

Acknowledgement

I would like to express my sincere gratitude to all those who have contributed to the completion of this research.

First of all, I am greatly indebted to my principal supervisor, Dr. Helen Slatyer, for her valuable insight, clear guidance and generous assistance. Dr. Slatyer has patiently guided me through all the stages of this research with constant encouragement and inspiration. Her feedback has been valuable in helping me in starting to think and write like a researcher. I am also deeply grateful to my co-supervisor, Dr. Jean Cho, for her professional expertise, unfailing support and mentorship. Dr. Cho has devoted a considerable amount of her time to helping me understand my data. Also, her kind and clear advice has helped me overcome many difficulties during this journey.

I would like to give special thanks to Dr. Wonbo Kim, who inspired me to become a researcher and set out on this journey. In addition, I would like to thank my research participants, without whom this research would not have been possible.

I would especially like to thank my dearest friends and colleagues for their encouragement and support. Particularly, I would like to acknowledge Eunhye Ko, Sunjae Won, Christine Myung, Weiwei Zhang, Gyeyoung Lee, Xiaomin Zhang, Sunyoung Jin, Suzy Zenkins, Erin Pullukaran, Shubhankar Bera, Xuan Ha and Sophia Ra. To all the friends and colleagues whose names are not mentioned here, please know that it was only space limitations that prevented me from listing all of you.

Lastly, I would like to offer my heartfelt thanks to my family. I am truly grateful for their love and support that has always been the bedrock on which I have stood. I could not have come this far without them.

CHAPTER 1: INTRODUCTION

There has been a growing number of formal education programs for interpreters around the world since the 1990s (Gile, 2009). Those who wish to become an interpreter learn interpreting skills from the programs. Once interpreting students successfully complete the program and potentially gain certification, it is deemed that they are competent enough to work as professional interpreters. However, as Hoffman (1997) suggests, the completion of education is not the end point in developing interpreting skills. It is generally believed that interpreters need to continue to develop their interpreting skills throughout their careers to maintain their expertise and develop new skills and techniques. But once interpreting students complete their interpreter education program and start their professional careers as freelancers, they have far fewer opportunities to receive productive external feedback about the quality of their performance (Y. H. Lee, 2005; Motta, 2016; Slatyer, 2015). In addition, even during their education, the limited number of face-to-face hours means that interpreting students do not receive sufficient one on one feedback about their performance from their teachers (Hartley, Mason, Peng, & Perez, 2003). In this regard, the recent literature in interpreter education examines self-assessment as a useful learning skill that students should be equipped with in order to become autonomous and lifelong learners. The benefits of self-assessment have been identified in the literature in general education (e.g. Boud & Falchikov, 2007; Cassidy, 2007) and in specific domains of education such as in writing (Nielsen, 2014).

When I was an interpreting student, I carried out a small research project about interpreter students' perception of self-assessment as a part of the program requirement for my research methods class. The results of the project showed many of my colleagues were reluctant to self-assess their performance, or even listen to the recordings of their own performance. This

interesting result stimulated my interest in self-assessment and its instrumental role in interpreter education, and inspired me to investigate the current practice of self-assessment in interpreter education. I found that the area has been under-researched in the context of interpreter education despite its importance in learning being widely recognised. More specifically, little research has been carried out on interpreting students' ability to self-assess and the pedagogical implications of self-assessment. Understanding the current phenomenon in a natural learning environment where self-assessment is incorporated in an interpreting program might be a good starting point for studies in self-assessment in interpreter education.

This research, therefore, aims to explore the competence of interpreting students in assessing their own interpreting performance, the relationship between the accuracy of students' self-assessment and the results of their performance in the final exam, and lastly, the relationship between students' abilities in self-assessment and the instruction that they receive from their teachers. In an attempt to understand the students' behaviour in self-assessment in depth, I chose to employ a mixed methods approach that uses both quantitative and qualitative approaches to investigate this complex phenomenon (Dörnyei, 2007). Scores (quantitative data), journals and interviews with teaching staff (qualitative data) will be analysed to shed light on the self-assessment strategies adopted by interpreting students and how the processes of self-assessment are implemented in interpreter education in practice. It is hoped that by gaining a deeper understanding of the practice of self-assessment in interpreting classrooms, interpreter education can be enhanced in a way that will practically benefit interpreting students by equipping them with self-assessment skills.

This thesis consists of six chapters. Following this introductory chapter, Chapter 2 examines the literature to explore the role of self-assessment in the learning process and the need

for self-assessment in interpreter education. Chapter 3 considers the methodological approaches, the context of the study, and the methods for data collection and analysis employed in this study. Chapter 4 reports the results of the data analysis and Chapter 5 discusses the key findings in relation to each of my research questions, outlines the limitations of this study, and proposes some avenues for further research. Finally, Chapter 6 concludes with a summary of the findings and their implications for interpreter education.

CHAPTER 2: LITERATURE REVIEW

In this chapter, I firstly examine the literature on quality in interpreting performance that provides the theoretical foundation for the assessment of interpreters. A review of the skills required for professional interpreters follows in order to understand the requirements of formal education for interpreters. Drawing on the discussion of these two constructs, I provide the rationale for developing autonomous and self-regulated learning skills in interpreter education. Since interpreter education draws on the relevant theories of general education, I explore the literature on self-assessment in general education, which shows the importance and impact of self-assessment on learning. Finally, I explore the literature on self-assessment in interpreter education in order to better understand the notion of self-assessment and what we know about self-assessment.

2.1 Interpreting quality

The concept of interpreting quality has been a major focus of studies in the interpreting field and researchers have produced an abundant literature (Zwischenberger, 2010). Finding answers to questions like “What makes a good interpreter?” or “What is good interpreting?” is an essential foundation for assessing the quality of interpreting. The answers to these questions significantly impact not only professional interpreting practice but also on interpreter education, which ultimately aims to foster competent interpreters who can be accepted into the professional world of interpreting.

From the early years of interpreting research, the core criteria of interpreting quality have been associated with an equivalence relationship between the source and target discourse, leading to the concept of “good interpreting” being founded on “accuracy” and “fidelity” (Pöchhacker, 2016). Harris (1990, p. 118) explains the notion of accuracy as “re-express[ing] the

original speaker's ideas and the manner of expressing them as accurately as possible and without significant omissions." The English translation of Seleskovitch's (1968, p. 166, as cited in Pöchhacker 2016) wording "fidelité absolue" as "total accuracy" shows the two notions "accuracy" and "fidelity" can be used synonymously though the French word "fidelity" is usually translated as "faithfulness." This concept of interpreting quality led researchers to attempt to measure interpreting quality by quantifying errors in terms of omissions, additions and substitutions (see, for example, Barik, 1975/2002; Gerver, 1969/2002). However, the idea of measuring interpreting quality only by omissions, additions and substitutions, which is based on the assumption that those characteristics negatively affect the accuracy of interpreting, has been challenged by researchers. Hale (1997), for example, points out that "linguistic omissions and additions are often required to ensure accuracy" (p.211). Indeed, omissions and additions are often used as a strategy to achieve meaningful communication among interpreters, particularly to prevent miscommunication due to cultural differences (Napier, 2004; Slatyer & Chesher, 2007).

Since the 1980s, systematic investigations into interpreting quality have been carried out. Bühler's (1986) study is recognised as the first empirical research that sought to identify various factors that could affect interpreting quality. She surveyed International Association of Conference Interpreters (AIIC) members about sixteen criteria which, she proposed, are likely to be used for interpreting performance assessment. The criteria included linguistic and extra-linguistic factors such as "logical cohesion of utterance" and "pleasant voice." It should be noted that her criteria were developed for conference interpreting – in fact, most studies on quality are focussed on conference interpreting in the simultaneous mode, and this calls for more rigorous studies on other interpreting modes for other contexts. Nonetheless, her list of criteria inspired other quality survey studies (e.g. Chiaro & Nocella, 2004; Pöchhacker, 2012; Zwischenberger,

2010). These survey studies show the relative importance that interpreting professionals place on different quality criteria. Although they use different terms, the findings show similar results: that conference interpreters prefer content-related quality like “sense consistency” or “logical cohesion” over delivery-related quality such as “pleasant voice” or “grammar.”

Interpreting quality has also been investigated in terms of users’ expectation (e.g. Kurz, 2001). The biggest difference in perceptions of interpreting quality between professional interpreters and users lies in the question of accuracy. Users tend to judge the interpreting service they receive by the fluency of the interpreting. It is obvious that users who do not know both languages cannot judge the accuracy of the performance the way interpreting professionals can. However, considering how important it is for users to be satisfied with interpreting services, it is reasonable to attribute as much emphasis to the fluency of interpreting as to accuracy.

These research findings can serve as the foundation for interpreting quality assessment by informing the development of assessment criteria. Indeed, accuracy and fluency have been identified as the core concepts in assessment criteria suggested by researchers for interpreter education (see, for example, Hartley et al., 2003; J. Lee, 2008; S.-B. Lee, 2015; Riccardi, 2002). Liu, Chang, and Wu’s study (2008), which compared exit exams in eleven interpreter education programs in Taiwan, China, the UK and the U.S, shows that those concepts were widely accepted for interpreting assessment.

On the other hand, the concept of quality in the literature has not yet been translated into a set of concrete constructs; rather, it has been broken into a set of labels without clear definitions, which causes difficulties when they are used for measuring quality (Pöchhacker, 2015; Sawyer, 2004). After a thorough review of the literature, Fernández (2013) revealed the ambiguity of existing quality criteria. She particularly focusses on fluency and voice quality, and

shows how multiple denominations and definitions are used in the literature and in assessment materials. In the case of fluency, for example, she identifies two definitions: “One meaning is close to general proficiency in language, and the other is a more specialised sense, related to the temporal, suprasegmental features of speech, such as speech rate, uninterrupted runs of speech, number and duration of pauses (filled or unfilled), etc.” (p.55, italics in the original).

In addition to the ambiguity of the quality criteria in use, it is also recognised that assessment of interpreting is heavily reliant on experts’ holistic and subjective judgements even when a set of criteria is provided (Liu et al., 2008), resulting in low levels of agreement by raters (Sawyer, 2004). This rater behaviour might be the result of examiners applying their own standards based on their professional experiences and backgrounds, which might affect their perceptions about the relative importance of criteria (F. Wu, 2013). This problem could be resolved by rater training, but in practice, such training rarely takes place (Pöchhacker, 2015).

2.2 Skills required for interpreters

Interpreting is a complex task. While interpreting assignments are often undertaken by volunteers who may have not received formal training (Baker & Maier, 2011), professional level interpreting requires a set of highly advanced skills. To start with, as interpreting enables interlingual communication, it is evident that interpreters need to have high levels of proficiency in both of the languages involved in the communication (Pöchhacker, 2016). It is required that interpreters achieve some level of comprehension beyond simply recognising words and linguistic structures, because one word or linguistic structure in a source language may seem to correspond to a target language but mean something else in the target language (Gile, 2009). This means that interpreters should be able to comprehend a speech or discourse by focussing on its meaning and use linguistic skills such as discourse cohesion, paraphrasing, summarising and

identifying main ideas to convey the message (Bontempo, 2012). Gile (2009) adds that high-level interpreters must have good knowledge of, and ability to use, specific registers in a range of specialised areas (e.g. in law, science, politics, literary and other cultural areas). In addition to a thorough knowledge of the languages used in the interpreting task, in order to be able to interpret in a professional manner, interpreters need to be equipped with technical skills such as note-taking, the ability to work in different modes (e.g. simultaneous, consecutive, sight translation), the ability to manage the social organisation of communication (e.g. turn-taking and seating arrangements), the ability to use long- and short-term memory, and the ability to make complex choices under pressure (Hale, 2007). In conference interpreting (usually using the simultaneous or long consecutive mode), interpreters also need to work under time pressure that requires them to respond to spoken language in real time or very rapidly, a skill which is not acquired through foreign language learning (Gile, 2009). While interpreters require these linguistic and technical skills, coping strategies to deal with emotional stress are also critical. Interpreters can face situations in which they need to manage with mainly negative or extremely emotional content and personal stressors that may impact their well-being and performance. Therefore, interpreters should have the ability and knowledge to employ coping strategies to deal with occupational stress (Bontempo, 2012).

2.3 Interpreter education

The number of formal interpreter education programs has increased with the primary goal of developing high-level professional interpreting skills (Pöchhacker, 2016). Many education programs still follow traditional models of interpreter education, which can be characterised as a "master-apprentice" type of learning (Moser-Mercer, 2008; Sawyer, 2004). This teaching approach was established by the very first generation of teachers of interpreting, who were

accomplished professional interpreters (Pöchhacker, 2016). It most likely takes the style of an apprenticeship, where teachers transfer their know-how and professional knowledge to students by giving them real-life tasks with which to practice (Pöchhacker, 2016), and corrections or comments on the student's performance are delivered verbally in classes (Gile, 2009).

Through repeated intense practice, observation, and corrective feedback, students automatise their use of the skills they learn from their teachers. According to Gile's (2009) Effort Model, the automatised skills require less mental energy, the total capacity of which is limited. In other words, it becomes less effortful to carry out interpreting processes when skills are automatised. This approach to interpreter education is likely to foster routine expertise (Moser-Mercer, 2008). Routine experts are highly competent in their domain, with an outstanding level of accuracy, speed and skill automatisation, and they are able to enhance their skills to perform more efficiently by establishing patterns and processes (Hatano & Inagaki, 1986).

However, the traditional apprenticeship model of pedagogy has limitations due to the nature of interpreting tasks and skill acquisition. Firstly, although automatised skills are desirable, routine experts' problem-solving approach is limited to them being able to use their skills in a familiar way (Hatano & Inagaki, 1986). Considering that interpreters often encounter new, challenging working conditions (and/or situations), interpreters should be trained as adaptive experts, who are capable of adjusting their skills flexibly and swiftly (Moser-Mercer, 2008). Secondly, the pedagogical approach might make interpreting students dependent on teacher feedback, while at the same time individual students often do not receive thorough feedback on their performance from their teachers (Choi, 2004) due to the limited number of training hours with their teachers (Ivars & Calatayud, 2013).

In addition, the duration of most interpreter education programs is only between six months to two years. According to Hoffmans' (1997) categories of the levels of expertise in interpreting, the expected level of expertise of those who successfully complete their interpreting education is the "journeyman" (p.199). The journeyman is at the midpoint of Hoffmans' categories, which describe seven levels of expertise (naïve, novice, initiate, apprentice, journeyman, expert and master). It is therefore clear that interpreters need to continue advancing their interpreting skills after completing formal training in order to achieve the expert and master levels. The idea of continuous learning post-education is not surprising considering that in most domains of expertise a minimum of ten years of intense practice is required (Ericsson, 2004; Ericsson, Krampe, & Tesch-Romer, 1993). But the problem is that interpreters rarely receive productive feedback about the quality of their interpreting performance once they enter the workforce as professionals (Y. H. Lee, 2011). This is largely because interpreters usually work as freelancers (Y. H. Lee, 2011; Slatyer, 2015).

Recognising these challenges, there is a strong argument in the literature that interpreter education should include the acquisition of skills that are necessary for interpreters to become adaptive, autonomous, and lifelong learners. Gile (2009) advocates incorporating theory in interpreter education (where the focus is essentially on practice) because he believes that theoretical components help interpreting students better understand phenomena, difficulties, strategies and tactics in professional practice. For example, theories help students understand why interpreters experience difficulties in understanding simple sentences, why speech sentences can be ungrammatical, and why interpreters should be careful not to heavily rely on their notes in consecutive interpreting. He explains that such understanding enables students to use appropriate strategies and tactics during and after their training. In addition to teaching interpreting theories,

self-regulated learning, which is a core conceptual framework of learning in the field of educational psychology (Panadero, 2017), has gained attention as it is recognised as a prerequisite for autonomous, lifelong learning and expertise acquisition (Hild, 2014).

2.4 Self-regulated learning

Theories of self-regulated learning have evolved in the field of educational psychology, where scholars in the field have paid attention to the impact of individual differences in learning since the early twentieth century (Zimmerman, 2002). Scholars in the field became interested in what leads some students to become highly motivated and learn quickly while others experience difficulties in understanding what they learn and loose interest easily (Zimmerman, 2002).

Introducing the term "metacognition," Flavell (1976) suggests that it is metacognition which makes the difference in cognitive development. Metacognition refers to one's knowledge about one's own cognitive processes and/or products, as well as one's monitoring and regulating of these processes and/or products (A. Brown, 1987; Flavell, 1979; Jacobs & Paris, 1987). Through metacognitive monitoring, learners continually evaluate their standards and learning strategies to determine whether they are suitable for their learning, so that they can make necessary changes using the results of their monitoring (Winne, 2017). Learners who have a higher level of metacognitive skill in learning show better academic achievement because they are aware of *what* they know and *why and how* a strategy works or not (Silver, 2011).

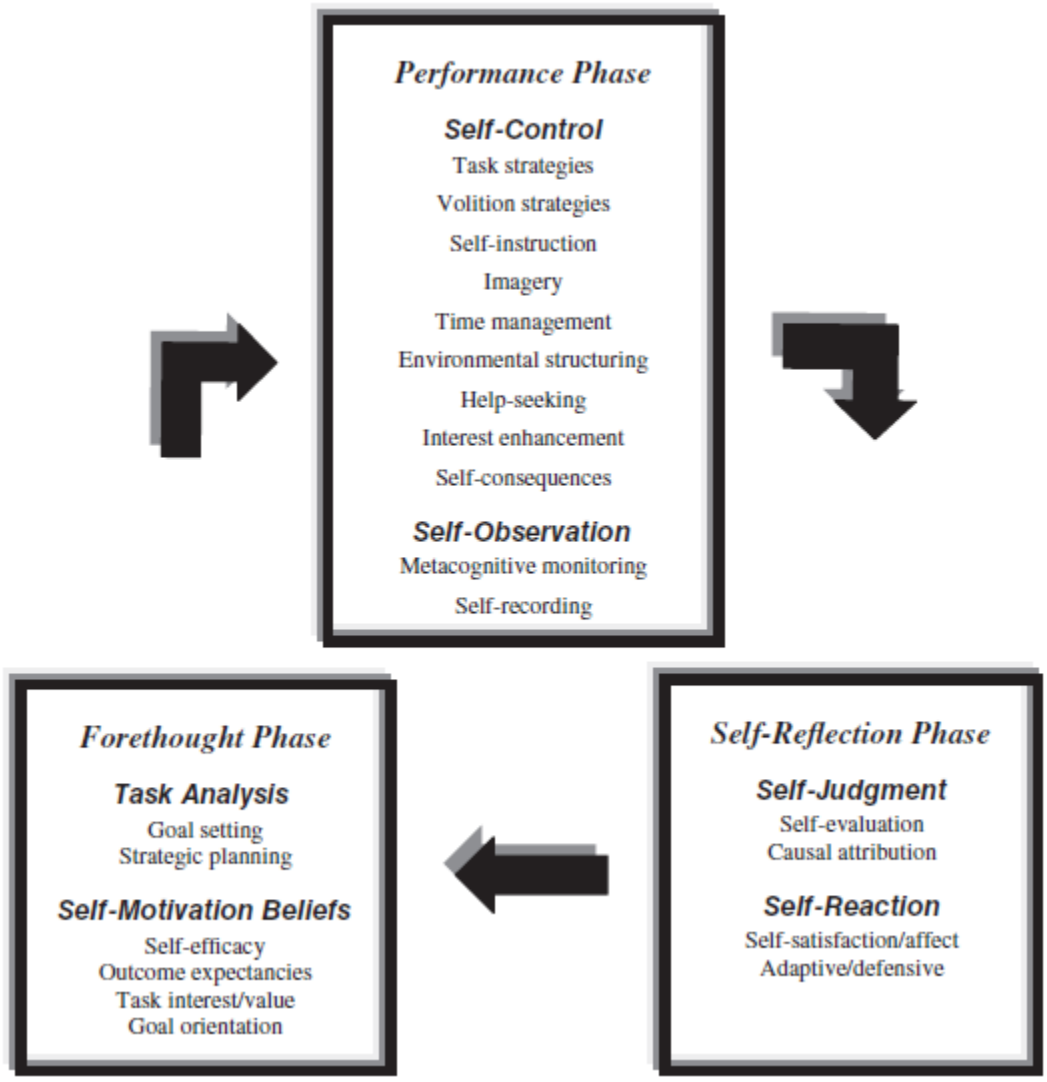
Metacognition is the predominant process of Winne and Hadwin's (1998; Winne, 2011) self-regulated learning model, which unfolds over sequential and recursive phases (Winne, 2011). In phase 1, learners develop their understanding of a task. They can use their metacognitive knowledge; for example, they ask themselves how much they know about the topic or if they need to seek more information for the task. In phase 2, learners set goals and plan

for the task. Metacognition plays a role in choosing study tactics by enabling learners to use their knowledge of when, why and how they can use them. In phase 3, learners carry out the task using the chosen tactics and monitor metacognitively to check how they are doing and if the outcomes of their task are likely to meet the goals they set in phase 2. In phase 4, learners evaluate the entire process of the completed task. They gain metacognitive knowledge about what they can or cannot do in the evaluation process. The knowledge gained from the process can be used to avoid making the same mistake in the future. This model suggests that metacognitive skills are advantageous for becoming a self-regulated learner.

One of the first scholars to develop self-regulated learning models, Zimmerman, points out that there are students who are able to use, and are aware of the importance of, metacognitive strategies in learning, but do not make use of the strategies (Zimmerman, 2002). He suggests that motivation additionally plays a major part in self-regulation, saying that metacognition explains how students regulate their learning, and motivation explains why they engage in self-regulated learning. Reflecting this idea, motivational variables are integrated into Zimmerman's (Zimmerman & Moylan, 2009) self-regulated learning model, with metacognition in three cyclical phases: forethought, performance, and self-reflection (Zimmerman & Schunk, 2011) (see Figure 1). During the forethought phase, students analyse their tasks, set goals and select strategies – metacognitive processes similar to phases 1 and 2 of Winne and Hadwin's (1998; Winne, 2011) model. What is additionally considered in the forethought phase is that students' motivational feelings and beliefs influence the metacognition. During the performance phase, the students carry out the task using metacognitive monitoring strategies to observe their progression while controlling various strategies to keep themselves motivated. During the self-reflection phase, the students evaluate their performances against a standard or their goals, attribute their

results and gain perceptions of satisfaction. The result of this phase is particularly important in this model since it influences the forethought phase of the next task. For example, students who are satisfied with their results show increased motivational beliefs in the next forethought phase (Zimmerman & Kitsantas, 1999). Corroborating this, a meta-review analysis (Dignath, Büttner, & Langfeldt, 2008) on self-regulation intervention studies reveals that training students' self-reflective strategies has a significant impact on both motivation and academic achievement.

Figure 1 Phases and subprocesses of self-regulation



(Zimmerman & Campillo, 2003)

Depending on the level of ability in self-regulation, different characteristics are shown among self-regulated learners (Zimmerman, 2002). Novices in self-regulated learning set goals that are usually too general to plan specific learning strategies and monitor and evaluate progress or outcomes. In particular, they tend to compare with others who are also in the process. This makes it difficult for them to see their own progress, and thus they attribute their failure to their low ability for the task. On the other hand, expert self-regulated learners set specific goals and strategies accordingly. As a result, they evaluate their performance against their goals or

appropriate standards, rather than against the performance of others. Also, they tend to attribute their failures to strategies or efforts. These characteristics of experts lead to more satisfaction and thus greater motivation in their learning (Zimmerman, 2002).

As discussed, these two models show how effective self-regulated learning takes place. Although there is the difference that Zimmerman's model considers the emotional aspect of learners, both models share the metacognitive aspects of self-regulated learning. Reflecting the entire learning process through self-assessment plays a key role in both models (phase 4 in Winnie and Hadwin's, and self-reflection phase in Zimmerman's model), as it occurs in the last phase of each model and thus keeps learners engaged in self-regulated learning.

2.5 Reflective practice

In professional education, reflective practice is commonly incorporated as a key component of the curriculum, such as in teacher education (Larrivee, 2000) or nursing education (Johns & Freshwater, 2009). Influenced by Dewey's theory of experience, Schön (1983) highlights the importance of reflection in learning, and coined the term "reflective practice." He argues that theory and practice can be closely aligned by reflective learning in professional education. In his view, there are two different types of reflection: reflection-in-action and reflection-on-action. Reflection-in-action is thinking on the event while it is taking place, often described as "thinking on your feet," whereas reflection-on-action is retrospective thinking about the event and analysing your practices to explore the reasons for the event and what could have happened differently (Schön, 1987). Another scholar influenced by Dewey's work, Boud (1995) agrees that effective learning comes from reflection, and stresses that for effective learning, the link between learning and reflection should be strengthened as it does not occur naturally. A number of professional education studies, particularly novice-expert studies, show reflection is

one of the distinctive characteristics of experts. For example, a study carried out by Tsui (2009) identifies reflection as one of the distinctive qualities of experts and experienced "non-experts" in teacher education. The study compared four professional ESL teachers, including one novice teacher's and three experienced teachers' teaching styles. Its results showed that among teachers with similar work experience, only the teacher who was recognised as an expert was constantly engaging in reflective practice. Also, experienced "non-expert" teachers' teaching styles were not significantly distinguished from the novice teacher.

Gibbs' (1988) reflective cycle (Figure 2), a classic model of reflection, demonstrates how an experience becomes a meaningful learning activity through conscious reflection.

Figure 2 Gibbs' reflective cycle

(Gibbs, 1998)

A number of other models of reflection have been developed, such as Kolb's experiential learning cycle (1984) and Boud, Keogh and Walker's reflective process (1985). Another, Johns'

Model of Structured Reflection (1994), is often used in the healthcare field (Finlay, 2008). While these models offer differing descriptions, Quinn (2000) identifies three fundamental processes common to models of reflection:

- Retrospection: i.e. thinking back about a situation or experience
- Self-evaluation: i.e. critically analysing and evaluating the actions and feelings associated with the experience, using theoretical perspectives
- Reorientation: i.e. using the results of self-evaluation to influence future approaches to similar situations or experience. (Quinn, p.81)

Self-evaluation (or self-assessment) is a common component in reflective learning models, playing a critical role in constructing meaning from experience. For effective learning through reflection, self-assessment must be involved in the process, and not simply as a component in reflective learning and self-regulated learning: the benefit of self-assessment in learning has been recognised in education, and self-assessment has evolved significantly in the framework of assessment.

2.6 Assessment[1] in education

The primary purpose of assessment in education is to determine whether students are qualified to move on to the next phase of education or receive certification (Wiliam, 2000). For these purposes, assessment usually takes place after a certain period of learning to examine whether students have gained the knowledge or skills they are supposed to have gained. The

---

[1] Assessment or evaluation

The terms “assessment” and “evaluation” are often used synonymously in the education literature (Taras, 2005).

results are the final outcomes of their study, and cannot be improved, and thus they fear making errors and showing their misunderstandings (Earl, 2005). In high-stakes assessments, which can change the course of a student's education, career, or life, it is likely that students become strategic in taking the test rather than focusing on learning, qualified as "negative backwash effect" (Biggs & Tang, 2011, p. 197). In addition, strategic test taking such as this does not provide an opportunity for students to use the results for their learning because students do not receive specific information about what they have or have not learned (Earl, 2005), and the results are released after the learning event is concluded (Biggs & Tang, 2011). Assessments used for this function are effectively separated from learning.

In contrast to assessments leading to negative backwash, educational assessment can be used to help students learn (Black & Wiliam, 1998; Boud, 1995; Wiliam, 2011). Recognising the need to distinguish the different purposes of assessment, Bloom and his colleagues (1971) suggest using the terms "summative assessment" and "formative assessment," an extended concept of the terms employed by Scriven (1967), who uses them only in the context of curriculum evaluation, where a curriculum is assessed for its appropriateness. According to Bloom's definition, the assessment described at the beginning of this section is a summative assessment, focussing on an outcome of learning, whereas formative assessment emphasises improvement over the course of learning. This definition makes it explicit that assessment can be useful for teachers and students to improve their teaching and learning. To be formative, assessment should produce information, or "feedback," that identifies a gap (Sadler, 1989) between the current level and the desired level (Ambrose, Bridges, DiPietro, Lovett, & Norman, 2010). This means students need to reveal their actual knowledge or ability to perform an assessment task. Therefore, in formative assessment, students need to feel that it is okay to make

errors to prevent the negative backwash effect (Biggs & Tang, 2011). Formative feedback allows teachers to notice where students are having difficulties and make the necessary adjustments to their instruction in order to raise students' achievements (Boston, 2002). Likewise, students also use the feedback to decide what to do to make improvements (Chappuis & Chappuis, 2007). Feedback can be used to promote learning if it is provided when there is time to make the adjustment. Therefore, formative assessment usually takes place during instruction (Boston, 2002; Chappuis & Chappuis, 2007; Evans, Zeun, & Stanier, 2014).

These differences between summative assessment and formative assessment lead to the problem that each purpose of assessment can be served only by a different assessment system. In the current school system, which eventually requires summative assessment, this tension has impeded the application of formative assessment in class, because teachers are already overloaded in the provision of formative feedback (Black, Harrison, Lee, Marshall, & Wiliam, 2004; Gibbs & Simpson, 2004). Teachers are under pressure to carry out an assessment with a different form more frequently for formative purposes. It is true that some assessments are better suited to summative use rather than formative use (Chappuis & Chappuis, 2007). For instance, a norm-referenced assessment that compares students' competence is effective if the purpose of the assessment is to select some of the students by quota, which is a summative use. For formative assessment, criterion-referenced assessments that give concrete ideas about how to improve are adequate (Wiliam, 2000).

Although the tension between summative and formative assessments produces some dilemmas in practice, the notion that formative assessment promotes student achievement is receiving support and recognition as a valuable learning tool (e.g. Gibbs, 2010). Emerging research has attempted to explore the positive link between formative assessment and learning

(Wiliam, Lee, Harrison, & Black, 2004; Wininger, 2005). For instance, Dunn and Mulvenon (2009) reviewed more recent studies on formative assessment interventions to investigate the impact of formative assessment on learning gains, and found that students' achievements improved in the studies.

With the growing interest in formative assessment in education, many definitions and activities have been offered (Black & Wiliam, 2009). Apart from the fundamental definition suggested by Bloom and his colleagues, another distinguishing aspect of formative assessment is that it includes learners' engagement in the learning and teaching process. The framework (Figure 3) that Wiliam and Thompson (2007) suggest consists of five key strategies for formative assessment that have been identified in the literature. It shows that students play an active role in their learning in the framework of formative assessment.

Figure 3 Aspects of formative assessment

<table>
<tr><th></th><th>Where the learner is going</th><th>Where the learner is right now</th><th>How to get there</th></tr>
<tr><td>Teacher</td><td>1 Clarifying learning intention and criteria for success</td><td>2 Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding</td><td>3 Providing feedback that moves learners forward</td></tr>
<tr><td>Peer</td><td>Understanding and sharing learning intentions and criteria for success</td><td colspan="2">4 Activating students as instructional resources for one another</td></tr>
<tr><td>Learner</td><td>Understanding learning intentions and criteria for success</td><td colspan="2">5 Activating students as the owners of their own learning</td></tr>
</table>

(Wiliam & Thompson, 2007)

The fifth strategy in this framework suggests that students are encouraged to actively engage in their own learning in the environment, as set by the teacher. Self-assessment has been identified as one of main class activities for formative assessment (see Sadler, 1989) since ideas about practical applications of formative assessment have been discussed without a theoretical basis

(Black & Wiliam, 2009). Providing a comprehensive theoretical framework for formative assessment, Black and Wiliam (2009) suggest that self-assessment can be used for the fifth strategy in the framework, recognising its relationship with learner autonomy.

As the concept of formative assessment evolves in education, scholars have been discussing its impact on students' learning and the practical application of self-assessment in class.

2.7 Self-assessment in education

As discussed in earlier sections, self-assessment is a key component in reflective learning and self-regulated learning, which conceptualise how students learn from experience and how students drive their own learning. It is recognised as a type of formative assessment. What, then, constitutes self-assessment, and how is it associated with learning?

Self-assessment has gained attention with moves towards fostering autonomous leaners (Lew, Alwis, & Schmidt, 2010). It is practiced in a variety of forms, ranging from simply asking students to grade their own work to involving them in providing formative feedback and developing explicit criteria (Panadero, Jonsson, & Botella, 2017). Taras (2010) classified the simple form as one of the "weaker models of self-assessment," that is a preliminary step to implementing self-assessment in practice. Falchikov and Boud (1989) provide a general definition of self-assessment as "the involvement of learners in making a judgement about their achievements and the outcomes of their learning" (p. 529). In his later work, Boud (1995) emphasises students' involvement in "identifying standards and/or criteria to apply to their work" in self-assessment (p.12). According to Boud, the crucial aspect of self-assessment is that students hold an understanding of what makes work good so they can make a quality judgement about their own work. Sadler (1989) agrees with his view, stating that a condition for

improvement is that students come to possess a similar concept of quality as that held by their teacher, and are thus able to judge their own work. He argues that it is students who take action for improvement, but if they merely follow the teachers' diagnostic feedback automatically, without understanding its purpose, they cannot improve. In this sense, students' self-assessment is "not an interesting option or luxury" (Black & Wiliam, 1998, p. 55).

By making a quality judgement of their own work, students formulate feedback for themselves (Andrade & Du, 2007). In other words, students become a source of formative feedback on their own work by engaging in self-assessment. This means that self-assessment could be a solution to a common issue of formative assessment, namely that feedback from teachers is insufficient for learning (McDonald & Boud, 2003). In addition, students become aware of their goals, identify their strengths and weaknesses, and check their progress in the process of self-assessment (Andrade & Valtcheva, 2009). These aspects suggest that self-assessment enhances students' ownership of learning, which makes self-assessment a core component of self-regulated learning (G. T. L. Brown & Harris, 2013; Schunk, 2003; Siegesmund, 2017).

The positive impact of self-assessment on learning has been investigated by many researchers in education. For example, Ross, Hogaboam-Gray, and Rolheiser (2002) examined students in grade 5-6 maths class (n = 259 treatment, 257 control) and found students who were taught to self-assess outperformed control students. The effect size was .40 which means "a student at the 50th percentile in the control group would have performed at the 66th percentile if he or she had been in the treatment group" (p. 53). Andrade, Du, and Wang (2008) investigated the effect of self-assessment on grade 3 and 4 students' writing ability. Their study showed that the students who were guided to self-assess their writing produced more effective writing and

received higher scores than those in the control group. After a meta-analysis of 24 studies on self-assessment implementation, G. T. L. Brown and Harris (2013) found that the median effect size range was from 0.40 to 0.45 and concluded that self-assessment seems to result in an improvement in students' performance.

Despite the positive impact of self-assessment on learning, some research suggests students are often reluctant to use self-assessment. Leach (2012) explores whether students opt to self-assess when it is optional in their course. She found that the majority of students (62%, 292 out of 472) chose not to self-assess. This result echoes a finding of Andrade and Du's (2007) study. In interviews, students reported that they did not apply self-assessment practice in other classes where it was not required, citing a lack of motivation and support for the practice. Another study conducted by Maguire, Evans, and Dyas (2001) showed that first-year university students were sceptical about self-assessment and to them, it was a "mechanical" task. They found that the students took a strategic approach to a self-assessment task to receive a good score for the task by investing minimal effort in it. Students in Olina and Sullivan's (2004) study reported that they put little effort into self-assessment because they found it difficult and they could not objectively judge their performance, indicating a lack of confidence.

Student's ability to self-assess has been discussed in many studies. Usually, researchers compare grades given by the student with those of the examiner. For instance, Austin and Gregory (2007) compared self-assessment by senior-level bachelor of science pharmacy students to external assessment (patients, instructors, and peers). Overall, the students overestimated their skills compared to their actual levels. On close inspection, those in the lowest actual percentile significantly overestimated their results while those who scored in the highest quartile slightly underestimated their results. Cassidy (2007) evaluated the accuracy of the self-assessments of

160 first-year undergraduate students. The comparison between student estimates and tutor marks shows there was no significant difference, and the two marks were positively correlated. Fifty-eight point five per cent of students estimated their mark within 10% of the tutor's mark, which demonstrates their capacity to accurately self-assess. On the other hand, Lew et al. (2010) study shows a different result. They examined the accuracy of self-assessment carried out by 3,358 first-year students in a post-secondary institution over a semester. The students' self-assessments were compared with peer and tutor-assessment. The analysis showed a weak to moderate accuracy in students' self-assessment abilities. As these studies demonstrate, studies about student's ability to self-assess show inconsistent findings, which indicates a need for more empirical studies in order to develop a deeper understanding of self-assessment.

Nonetheless, scholars recognise a variety of potential benefits of self-assessment in learning, confirming that self-assessment can enhance learning and academic achievement (G. T. L. Brown & Harris, 2013), lead to an increase in self-efficacy (Olina & Sullivan, 2004), promote the development of metacognitive engagement (Rivers, 2001), increase the effectiveness of self-regulated learning (Kostons, van Gog, & Paas, 2012) and enable students to become autonomous and lifelong learners (Boud, 1995; Cassidy, 2007; Taras, 2008). In this vein, self-assessment theorists argue that teachers should incorporate self-assessment into the curricula and encourage students to self-assess (G. T. L. Brown & Harris, 2013; McDonald & Boud, 2003). However, Panadero, Brown, and Strijbos (2016) specify that "to self-assess, one needs to learn self-assessment" (p.819). Like any other tasks, students cannot be expected to perform self-assessment with ease and accuracy (Panadero, Brown, and Strijbos, 2016). However, it also appears that many teachers lack the understanding to promote effective self-assessment in students and foster self-assessment in classes (Panadero & Alonso-Tapia, 2013).

Self-assessment theorists therefore recommend methods for implementing self-assessment in class. First, teachers should provide criteria to students. The criteria could be devised by the teacher, or co-developed by students and the teacher. When students are involved in the process of criteria development, they gain a better understanding of quality constructs thoroughly (Black et al., 2004; Nicol & Macfarlane‐Dick, 2006). Second, teachers should teach students how to apply the criteria by modelling (e.g. demonstrating self-assessment) and providing examples. Teaching how to apply the criteria is critical: scholars argue that we cannot expect students to know and use assessment criteria when they simply receive a rubric (Andrade & Valtcheva, 2009; Gibbs, 2010). Third, teachers should provide feedback on self-assessment and finally give students opportunities to practice self-assessment, as the acquisition of the skill does not occur at once (Panadero & Alonso-Tapia, 2013).

2.8 Self-assessment in interpreter education

In interpreter education, self-assessment is slowly gaining attention from researchers and trainers. Although it is as yet an under-researched area with only a few studies, the studies provide useful insight on the characteristics of self-assessment practices and how self-assessment can be incorporated into the interpreting classroom.

The literature recognises that self-assessment is a critical learning skill that needs the careful guidance of teacher and the development of clear practices. Choi (2004, 2006) proposes a metacognitive evaluation model whereby teacher and students work together in identifying strengths and weaknesses for improving interpreting skills. In the model, students evaluate their own performance against shared criteria and receive feedback from their teachers. According to Choi (2004; 2006), the model could effectively enable interpreting students to become confident in judging their performances by engaging them in self-assessment on a regular basis. Fowler

(2007) also advocates a role for teachers in developing self-assessment skills, observing that even experienced interpreting students are often unaware of the mistakes in their own performances, such as omissions. Therefore, she highlights that interpreting students need explicit criteria, guidance, opportunities for practice, and feedback provided by a teacher in order to learn how to self- and peer-assess. In her view, practicing peer-assessment is the first step to developing students' abilities in self-assessment, since they can first gain experience as assessors. These two papers, however, are conceptual, and lack empirical evidence demonstrating how self-assessment could be successfully developed within the proposed frameworks.

A small-scale case study conducted by Pinazo (2008) provides evidence to support the idea that interpreting students could develop self-assessment skills with systemic guidance. In his study, students were required to self-assess as a mandatory activity every time they practiced interpreting in class. The researcher provided a self-assessment sheet that required students to give themselves a score against the provided criteria. Throughout the academic year, teachers closely monitored individual students and provided advice by encouraging students to pay attention to their weaknesses. It was observed that students became engaged in their learning. For example, they consciously paid attention to their weaknesses and actively sought advice. At the end of the academic year, 90% of the thirty students who practiced self-assessment scored eight out of ten in their final assessments. This result indicates a positive effect of implementing self-assessment in an interpreting class. Pinazo reported that the students' self-assessment skills improved as they gained experience, although he did not explain how he measured the improvement.

Self-assessment skill is usually measured by comparing self-ratings with the ratings of experienced markers, often that of the teacher. This method has been employed in empirical

studies that have attempted to measure the accuracy of interpreting students' self-assessment. Two recent studies by Li (2018) and Han (In press), show that the accuracy of students' self-assessment increased over time. In Li's (2018) study, students performed self-assessment on sight translation tasks four times, with four-weeks intervals. It was observed that the gap in the mean values between the students' self-ratings and the ratings of teachers became closer with time. The first and the second times, the students overestimated their scores, but from the third time they were able to estimate their scores to a similar level as their actual scores, graded by their teacher. In Han's (in press) study, the students performed self-assessment three times with five-week intervals on consecutive interpreting tasks. Unlike Li's (2018) study, the students continued to overestimate their performance in each task. However, it was noticed that the accuracy of their self-assessments improved over time. Han (In press) analysed each assessment criterion (information completeness, fluency of delivery, and target text quality) for each language direction (English to Chinese and Chinese to English). The results of correlations improved in each criterion. For example, the Pearson's *r* value increased from 0.47 to 0.51 in "information completeness" for the Chinese-to-English direction. These two studies suggest that the accuracy of assessment could be an indicator of students' self-assessment abilities, and those abilities can be developed.

What should be noted about the two studies is that the researchers provided students with assessment criteria and explained in detail how to use the criteria. The students in Li's study were third-year undergraduates in a Translation and Interpreting program, and in Han's study the students were fourth-year undergraduates who had studied consecutive interpreting for three consecutive semesters, which indicates that they might be familiar with the qualities of a good interpreting performance. In Li's (2018) study, the teacher and the students developed the criteria

together, which means students had sufficient knowledge to contribute to devising the criteria. Yet, still, the teacher organised two sessions to guide students and answer questions about how to use the criteria.

It is important that students understand and are able to use assessment criteria for self-assessment (Fowler, 2007; Hartley et al., 2003; Y. H. Lee, 2005). The criteria help interpreting students to be conscious of their performance and thus be in control while performing, and also enable effective communication between students and teachers (Doğan, Arumi Ribas, & Mora-Rubio, 2009). However, some studies report that students encounter difficulties in using criteria due to confusion about the terms used (Hartley et al., 2003; Y. H. Lee, 2005). For example, in order to promote autonomous learning, Hartley et al. (2003) devised a set of criteria. The primary aim was to develop user-friendly criteria so that students would find it useful for their self-study. In the process of developing the criteria, Hartley and his colleagues investigated how well novice interpreting students and advanced interpreting students understood and explained nine commonly discussed components of interpreting quality criteria. The results suggest that as interpreting training progresses, interpreting students become more capable of distinguishing good interpreting, and describing interpreting performance with a clearer understanding of quality criteria, but with regard to “Fluency and Delivery,” there was no agreement in the description provided even by the advanced students. In addition, during a pilot test for the criteria’s user-friendliness, some students reported difficulties in distinguishing close concepts such as cohesion and logical links. This indicates that, for students to be able to use the metalanguage in assessment criteria, teachers’ skills and efforts are required to support the development of a common understanding (Lee, 2005; Dogan et al 2009).

In addition to the accuracy of self-assessment and the ability to apply criteria, another characteristic of interpreting student's self-assessment practice is that they tend to be very critical of their performance and focus on mistakes. In Bartłomiejczyk's (2006) study, 80% of remarks made by the students were negative. In a subsequent study (2007), she explicitly asked students to pay attention to both positive and negative aspects of their performance. Still, negative remarks were 20% more prevalent than positive remarks in their self-assessments.

CHAPTER 3: METHODOLOGY

The previous chapter examined the underlying constructs, theories and empirical studies regarding self-assessment in general education, and more specifically in interpreter education. The review shows the potential benefits of self-assessment in interpreter education, and indicates a need to develop a deeper understanding of how students develop self-assessment skills.

This study aims to explore interpreting students' self-assessment practices and to investigate the possible relationship between student self-assessment and teacher instruction in self-assessment through a case study in a context where self-assessment is implemented to meet certain stated learning outcomes. Reflecting the current practice in implementing self-assessments will provide insights for constructive teaching and better use of self-assessment. In order to achieve the aim of the study, three research questions were formulated:

RQ1: How competent are students in assessing their own performances?

RQ2: What is the relationship between students' ability to self-assess and interpreting performance outcomes?

RQ3: What is the relationship between teachers' instructions on self-assessment practice and students' ability in self-assessment?

The potential outcomes of the study could be a guide for teachers in directing students to be able to accurately assess their interpreting performance in view of supporting the reflective practice tasks that are required in their education and future practice.

In this chapter I present the underlying methodological approach to this study to answer the research questions and methods used for data collection and analysis.

3.1. Methodological approach

As we have seen in the literature on students' self-assessment competence, most studies have employed quantitative approaches to measure students' abilities to self-assess their performances based on assessment scores. By comparing students' self-ratings with teacher-ratings, we can see whether students are able to judge the quality of their interpreting performances using the same standards employed by their teachers. I have, similarly, adopted a measurement approach to answer research questions 1 and 2. A quantitative approach will allow me to compare the results of this study to the results of the empirical studies reported in the literature review. In addition, as the literature highlights, in order to be able to self-assess, the ability to use a set of criteria to make a judgement about the quality of an interpreting performance is essential. As a former student in the interpreting program where this study took place, I had experience in completing reflective journalling for the purpose of self-assessment and reflection during my studies. Therefore, I decided to use the journals as qualitative data to explore students' abilities in applying assessment criteria by analysing their journal entries.

For research question 3, I needed to obtain more detailed information about how the teachers instructed students to undertake self-assessment in their classes. Therefore, I chose to conduct one-on-one interviews with the teaching staff. The interview is a useful research instrument to understand the teachers' implementation of the self-assessment task and reflective practice in the classroom in more depth (Cohen, Manion, & Morrison, 2017). Depending on the researchers' knowledge or awareness about the research topic, or the purpose of the interview, the researcher can select the degree to which the interview is structured, ranging from unstructured to structured interviews (Cohen, Manion, & Morrison, 2017). The semi-structured interview is at the midpoint of the continuum that allows researchers greater flexibility than a

structured interview, but still allows control over the topic (Ayres, 2008). I went into the interviews with relatively concrete interview questions (in Appendix 4) about teachers' instruction, but I also attempted to understand the thoughts and reasons underlying their instruction by asking follow-up questions. Therefore, semi-structured interviews were the best option for this study.

Based on the above considerations, I decided to employ a mixed methods approach which incorporates both quantitative and qualitative data collection and analysis in a single study (Creswell & Clark, 2011). It is an emergent third paradigm of research methods that rejects the dichotomy between qualitative and quantitative research, particularly in human sciences (Tashakkori & Teddlie, 2010). The advocates of quantitative research approaches believe that only scientific methodologies can yield legitimate research outcomes that formulate laws. On the other hand, the advocates of qualitative approaches argue that realities are complex and cannot be simply generalised (Burns, 2000). In these purists' views, the two research paradigms are fundamentally different and thus incompatible (Johnson & Onwuegbuzie, 2004). However, mixed methodologists contend that researchers should move beyond the dichotomous view and research approaches to answer research questions. This view can be characterised as "methodological eclecticism" (Tashakkori & Teddlie, 2010, p.777). Johnson and Onwuegbuzie's (2004) statement clearly shows this perspective: "Mixed methods research also is an attempt to legitimate the use of multiple approaches in answering research questions, rather than restricting or constraining researcher's choices." (p.17)

In this sense, utilising mixed methods in this study allows us to understand the interpreting students' self-assessment ability from the perspective of the outcome (score) and the process (the journal), as well as its relationship to teacher instruction. (Creswell, 2014, p. 535)

3.2. Context of the case study

This study was undertaken in the Graduate Diploma and Master's programs in translation and interpreting studies at an Australian university and its joint offshore program in Korea. As professional postgraduate programs, these courses focus significantly on the acquisition of professional competence in translation and interpreting based on the study of professional skills. These professional skills are underpinned by theoretical study and informed by empirical research. Students are also expected to acquire a range of capabilities that prepare them for professional life, including autonomy and self-directedness. In this sense, self-assessment has been embedded in the program and individual course unit learning outcomes. The example below shows the relevant part of the intended learning outcomes of the program:

> **Learning Outcome 9:** regularly conduct self and peer reflection on interpreting performance
>
> **Discipline-specific attributes** of graduates of the programs
> Able to critically evaluate own and peer's translation(s) and interpreting(s)
>
> a. Evaluation skills
> i. Self-evaluation
> ii. Peer assessment
> iii. Self-editing and revising
> b. Strategies for
> i. Peer review
> ii. Critical analysis
> iii. Self-reflective skills
> iv. Self-correction
> c. Error analysis
> d. Diagnostic skills

The learning outcomes for each unit of study are aligned with the program learning outcomes, and are embedded in the assessment tasks. An example is shown below:

> By the end of semester, students should expect to be able to achieve the following knowledge/skills/techniques. Students should:
> […]
> 8. Demonstrate reflective and critical practice to become life-long learners

> The graduate capabilities for this unit are:
> - Critical, analytical and integrative thinking
> […]
> - Capable of professional and personal judgement and initiative

The program offers three levels of practical interpreting classes to scaffold the development of the skills required for professional interpreting practice: Interpreting practice 1 (IP1), Interpreting practice 2 (IP2), and Interpreting practice 3 (IP3). IP1 is the foundational level of interpreting, where students learn the basic skills of interpreting in sight translation, dialogue mode, and the commencement of consecutive mode. IP2, the intermediate level of practical interpreting, focusses on developing the techniques for dialogue and consecutive interpreting. These first two units are core units, and compulsory for both the Graduate Diploma and Master's Degrees. IP3, the advanced level, focusses on long ("classic") consecutive and simultaneous interpreting. This unit is selective, and only the students who have successfully completed IP2 can undertake it. The offshore program only offers IP1 and IP2. Therefore, students who wish to graduate with a Master's Degree in the offshore program come to Australia and take one more semester to complete their study.

There is a convenor (coordinator) for each unit who is responsible for designing and administering the unit. Different lecturers and/or tutors teach classes for each language pair (e.g. Chinese/English, Korean/English, etc.). Unit convenors also decide on the intended learning outcomes for the unit in collaboration with the convenors of other units to ensure alignment of the curriculum across the program. Therefore, different language-specific tutorials within each unit share the same learning outcomes. One semester consists of 13 weeks of classes.

3.2.1 Self-assessment in the interpreting practice classes in the program

Interpreting practice classes incorporate self-assessment tasks as a component of a reflective journalling assignment that students write after the completion of an interpreting task. IP2 in the offshore program, for example, requires students to write a self-analysis report on their performance in a mock exam, which is expected to be submitted within one week of the mock exam. The self-analysis report accounts for 0.5% of the total unit grade while the interpreting performance is not assessed by their teacher. In IP3, after students take a mid-term exam on consecutive interpreting, they are required to write a reflective journal to analyse their performance. Usually the students are given about two weeks to complete and submit their report. The journals are the only assignment assessed in the unit other than students' interpreting performances. The mid-term exam results account for 30%, the reflective journal 20% and the final exam 50% of the final grade.

3.3. Methods and procedure

In the following section, I will present the methods and procedure I used for data collection and analysis, and outline the process of ethics approval. All research undertaken by students and staff at Macquarie University conducting research on human subjects is required to obtain ethical approval.

3.3.1 Ethics approval

Ethics approval to recruit participants for this study was granted by the Macquarie University Human Ethics Committee on 30th October 2017 with the reference number 5201700979 (see Appendix 1). Permission to recruit participants and collect data in the offshore program was obtained from the Director of the program prior to the submission of the ethics

application. Participation in the study was voluntary and only those who consented were included in the study. The participants' ID and classes were coded to ensure anonymity.

3.3.2 Participant recruitment

The participant groups in this study were students enrolled in IP2 or IP3 in the second semester of 2017, as well as the lecturers and tutors teaching their classes. Recruitment took place from 3 November 2017 to 20 November 2017.

Student participants

With prior approval from the lecturers, I visited each class of IP2 and IP3 in Australia and IP2 in Korea to recruit students for the study. I explained the purpose of the study, the data I would need for the study, and the data collection procedure to the students. Printed consent forms were distributed to the students at the beginning of the class. They were given time to read the consent form carefully and decide at the end of the three-hour class whether they would participate or not. To make sure they did not feel coerced, I collected all the consent forms, signed or not, so that the students who did not consent could not be identified by others in the class.

A total of 47 students (out of 89 potential participants) consented to participate in this research. The demographic information of the student participants is shown in Table 1.

Table 1 The demographic information of the student participants

Level	Language pair	The number of students in the class	The number of participants	The location of the program
IP3	Chinese and English	16	9	Australia
	Korean and English	13	11	
IP2	Chinese and English	29	7	
	Multilingual and English	10	7	
	Korean and English	21	13	Korea
	Total	89	47	

Teaching staff participants

Lecturers were contacted via e-mail or personally and invited to participate in the study. I explained the purpose of the study, the nature of the interview and the data collection procedure. All three staff teaching on the target units consented to participate in the study.

Claire[2] and Amber have been teaching interpreting since 2004 and Jessica has been teaching interpreting since 2014. They are all professional interpreters. Claire has nineteen years, Amber has eighteen years and Jessica has eight years of experience as professional interpreters. In terms of pedagogical background, Claire has a Master's degree in interpreter training and attended several short-term interpreter trainer programs. Amber and Jessica have not received any interpreter trainer education.

3.3.3 Data collection

Since this study aimed to investigate the current practices around self-assessment in interpreting practice classes, the study principally collected operational data from the classes and therefore did not allow for piloting of the research instruments. The interview questions were informally "piloted" with my supervisors. I was not able to do a more substantial pilot with the target population due to the very small number of staff available for the study. Only students' self-ratings were required for the purpose of this study, in addition to the existing data. The data were collected from November 2017 to January 2018.

1) Grades
    a. Mid-term exam (IP3) or in-class interpreting task (IP2): students' self-rated scores and teacher-rated scores

[2] Pseudonyms are used for the staff to ensure anonymity.

b. Final exam: teacher-rated scores

c. Journal grades

2) Students' reflective journals

3) Interviews with the lecturers

Data collection procedure

Self-rated scores from the students

The students were asked to write down their self-rating out of 100 on their mid-term exam on the consent form if they agreed to participate in the study. This process occurred after they had submitted their self-assessment reports and before they received their marks and feedback from their teachers.

Teacher's ratings and the journals from the lecturers

The lecturers sent the teacher-rated scores on the students' interpreting performances (the mid-term, in-class performance for IP2, and the final exam), the journals and the journal scores to the principal supervisor via email. Offshore student journals are in Korean, and all the journals contain references and examples in LOTE (Language other than English). Excerpts in LOTE used in this study were translated either by the author or the teaching staff who use the language.

Interviews with the lecturers

The lecturers were invited to one-on-one interviews. Interviews took place in a meeting room or teacher's office. Each interview lasted from 20 to 30 minutes. The interviews were audio recorded and transcribed. The lecturer in Korea chose to use Korean. The interview transcript was translated into English and double-checked by the author after analysing the transcripts in Korean.

3.3.4 Final data set

After collecting the data, it became apparent that the methods and procedures for interpreting students' self-assessment practice in the different classes were not standardised. IP2 Chinese-English and Multilingual-English classes used dialogue interpreting tasks for the in-class interpreting performance on which the self-assessment assignment was based. Also, these two IP2 classes used different scoring rubrics for self-assessment. I eliminated the data from these IP2 classes as they did not allow for any pattern identification, nor for comparison. The number of participants who had consented and whose data were eliminated was 14 (7 students of IP2 Chinese-English class and 7 students of IP2 Multilingual-English class).

Consequently, the data from the three classes that implemented the assessment of the consecutive interpreting mode and used the same rubric for the self-assessment assignment remained. These factors meant that there was greater consistency in the data set leading to more meaningful conclusions due to the larger number of samples that could be compared. The remaining classes were IP2 Korean-English (IP2Kr, n=13), IP3 Korean-English (IP3Kr, n=11), IP3 Chinese-English (IP3Ch, n=9).

3.4 Data analysis

The remaining data, following collection and the elimination process outlined above, were coded and cleaned for data analysis.

3.4.1 Data preparation

All student-related data (journals and ratings) were sent to the principal supervisor to ensure confidentiality. To preserve anonymity, the principal supervisor coded the students' ID and data, and I replaced the lecturers' names with pseudonyms and changed the codes of the classes.

During the data cleaning, I deleted the data of two students in IP3Ch who consented but did not provide their self-ratings. I organised the final data set of scores, which were the self-ratings and teacher's scores, in one table in Excel with class codes and student codes. Also, I converted the teachers' scores of IP3 classes into scores out of 100 since the original scores were rated out of 30 (the mid-term exam) or 50 (the final exam).

3.4.2 Score analysis

The scores were analysed quantitatively using the Statistical Package for the Social Sciences 25.0 (SPSS 25.0). Firstly, descriptive statistics for self-ratings and teacher-ratings on the midterm exam, and teacher-ratings on the final exam and the journal scores were calculated. Secondly, to determine if self-ratings were related to teacher-ratings on the mid-term exam for research question 1, the Pearson product-moment correlation was used. For research question 2, a multiple regression (ANOVA) was used to determine the degree of variance in the final exam scores (the outcomes) that can be explained by the accuracy of the self-assessment (the difference between the self-ratings and teacher-ratings) (Laerd Statistics, 2015). A multiple regression was used rather than simple linear regression to determine the relative contribution of accuracy combined with the journal score and the mid-term exam scores. It was assumed that if the accuracy has an impact on the final exam, there would be significance in the model.

3.4.3 Journal analysis: the coding process

The journal entries were qualitatively analysed as they provide insights about the students' self-assessment ability. This study adopted content analysis for a deep understanding of the journals and how the students went about their journalling. This analytic method involves sequences of coding, the recognition of patterns in the data, and the interpretation of the data

(Dörnyei, 2007). The analysis was undertaken using the NVivo 11 and 12[3] qualitative research software (QSR International, 2018). After de-identification, the journals were sent to my supervisor, who uploaded them to a shared Dropbox folder after de-identification before I imported them into NVivo for coding.

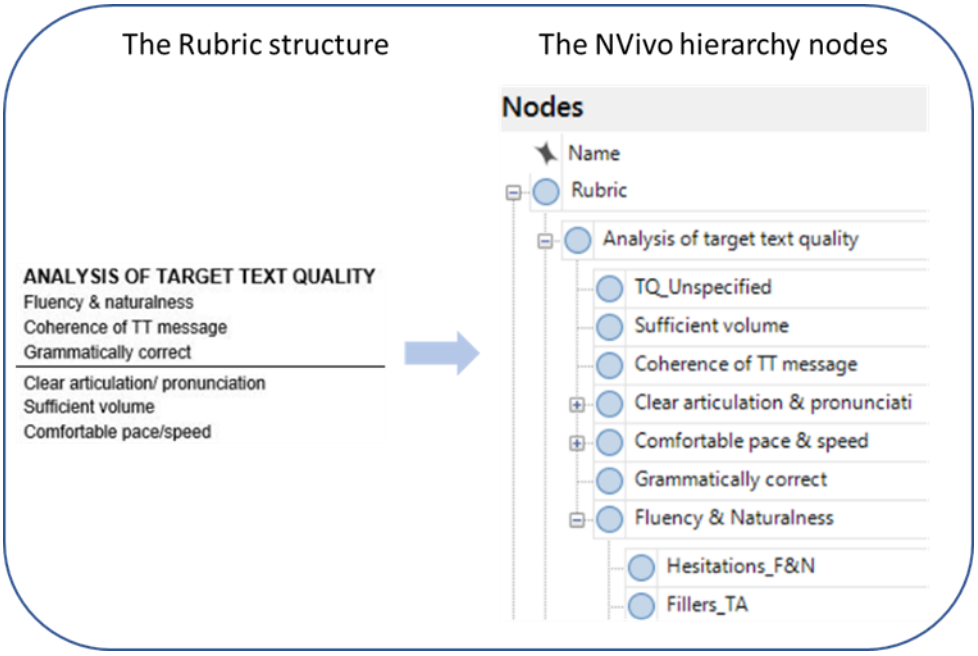
The coding process consisted of three steps. The first step was to familiarise myself with the content of the journals and plan the coding process; during this stage I read the journals multiple times to familiarise myself with the content and to identify potential challenges for coding. I then carried out two cycles of coding.

After familiarisation, I decided to use a combination of inductive and deductive approaches to analysis. Inductive coding starts with a close analysis of the data to generate ideas while deductive coding uses a set of pre-planned themes to be explored and to frame the analysis (Kaefer, Roper, & Sinha, 2015). During the first cycle of coding, I started with deductive coding using the rubric provided to the students for the self-assessment task. I created a hierarchy of nodes representing the structure of the rubric as shown in Figure 4 below. The full rubric is presented at the end of this chapter (Figure 6)

---

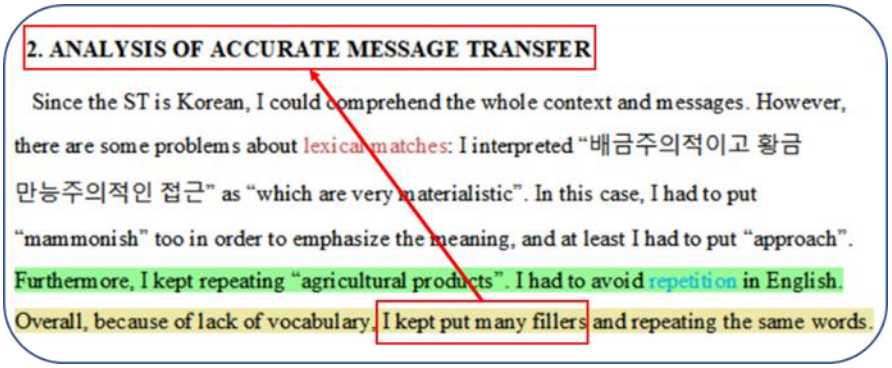
[3] NVivo was updated in the course of analysis

Figure 4 Deductive coding approaches



During this process, I also used an inductive coding approach. I noticed the students used specific terms that were not included in the rubric to discuss the criteria. For example, "fillers" were often used when the students discussed "fluency & naturalness" and "misinterpreting" was used to discuss "analysis of accurate message transfer." Therefore, using inductive coding for this type of references, I created a sub-category (a "child node" in NVivo terminology) under the criterion with which the students associated the term, as shown in Figure 5 below.

Figure 5 Inductive coding approaches



Similarly, in cases where a reference by a student in their journal did not relate to any of the criteria in the rubric, I created an "unspecified" node to which I assigned such references, as shown in the example below:

> "In the third paragraph, it seems to me that I failed to deliver the same meaning as the original text[4]." (IP3Kr12)

In addition, the rubric contains the criterion "comprehension of ST," which is the only process-oriented criterion, whereas other criteria are product-oriented. For these criteria, students discuss the quality of the interpreting output. Consequently, when responding to the criterion of "comprehension," students discussed the quality of their performance by referring to product-oriented criteria. In order to avoid one reference being counted multiple times, I decided not to make a separate node for "comprehension of ST."

During the second cycle of coding, based on my understanding of the students' writing styles and expression gained during the first cycle of coding, I categorised the references in "unspecified" into appropriate rubric nodes. If a reference did not match a specific sub-criterion, the reference was assigned to the broad category, "Target text quality" or "Accurate message transfer." Moreover, nodes that were created using the inductive coding approach were relocated under the appropriate node of a criterion where this became apparent. For example, the reference below was assigned to a node "self-repair" under "fluency & naturalness."

> "Though I tried to check the grammar and word choices before I speak, I corrected my words many times." (IP3Kr8)

For this process, I referred to the literature about interpreting quality assessment (e.g. Lee, 2006; Wu 2013; Baker 2011) and also consulted with lecturers who were experienced

---

[4] The students' comments are reported verbatim and have not been corrected for grammar or spelling.

interpreting examiners to make sure these references were appropriately coded. During the first and second cycle of coding, I repeatedly checked for consistency and used memos and annotations to help me review my interpretation of and decisions about the references.

Once the coding was completed, one of my colleagues was invited to independently check the nodes and assigned references to qualitatively check the reliability of my coding. I first explained to her how I assigned a node to a reference. Then we went through a sample of journals to ascertain whether she was assigning codes in the same way I did, and checked the consistency of my coding. When she raised a question about a reference, I checked, revised, and applied the new code across all the references in that particular node. Then she and I checked all the coding together.

After this process, I used the matrix coding function on NVivo 12 to explore the coded data. It showed the frequency of comments made by each student about each criterion. I extracted the table to Excel and calculated a percentage of each class's node frequency for each criterion to make charts for further analysis. For example, in the case of "omissions, additions, substitutions," using the sum of the frequency of comments about this criterion from all participants, the data was extracted from NVivo12 and imported to Excel. Excel automatically produced a chart showing the percentage of "omissions," "additions," and "substitutions" for each class.

3.4.4 Interview analysis

The audio-recorded interviews with the teaching staff were transcribed and the data was qualitatively analysed using content analysis to examine the responses to each of the questions. Since the number of interviewees and questions were small, the data was manually analysed,

using Microsoft Word 2016, rather than NVivo to identify commonalities and differences among the responses of the teaching staff in regard to their instruction in and approaches to self-assessment and the assignment. Transcriptions were uploaded to a new file and the range of responses for each question and each teacher were analysed. Summaries of the responses to key ideas were annotated using the comments tool.

Figure 6 The rubric provided to the students for the assignment

CONSECUTIVE INTERPRETING - REPORT MARKING SHEET

You are required to write a report analysing your interpreting performance, and the interpreting decisions that you made. You need to closely analyse your interpretations for 2 key areas: target text quality and accuracy of message transfer. Please include specific examples (min 1 per key area) of what you judge as unsuccessful target text choices or behaviours, and discuss improved decisions you could make in future. Make sure you respond to the assessment criteria below, as these competencies are what you will be marked for. (1,200 – 1,500 words)

SELF-ASSESSMENT REPORT CRITERIA	%
ANALYSIS OF TARGET TEXT QUALITY Fluency & naturalness Coherence of TT message Grammatically correct	/30
Clear articulation/ pronunciation Sufficient volume Comfortable pace/speed	
ANALYSIS OF ACCURATE MESSAGE TRANSFER Comprehension of ST Lexical/ conceptual matches Omissions/ additions/ substitutions?	/30
Equivalent register Equivalent affect	
DISCUSSION OF FUTURE IMPROVEMENTS (minimum 1 per section above)	/20
CLEAR STRUCTURE & GRAMMATICALLY CORRECT Extra marks for references to literature (incl. APA referencing)	/20

GRADE: **/100** **MARKER:**

CHAPTER 4: RESULTS

In this chapter I present the results of the study. First, I present the results of the quantitative analysis of scores, followed by the qualitative analysis of the journal entries. Finally, I present the results of the one-on-one, semi-structured interviews with the lecturers.

The analysis of scores consisted of descriptive statistics, the correlation between self-rating and teacher-rating on mid-term exam using Pearson product-moment correlation, and the relationship between self-assessment and performance outcomes using ANOVA. In reporting the results of the journal analysis, I present frequency counts of students' references to each of the criteria of the rubric in their journals. Lastly, I present the results of the content analysis of teacher interviews with the teachers' questions and responses.

4.1 The results of quantitative analysis

Quantitative data from IP3Ch, IP3Kr and IP2Kr were analysed using SPSS, as outlined in the previous chapter. Following the elimination of the non-conforming class assessments, the remaining data formed a very small data set for the purpose of quantitative analysis. This limited the range of analyses that I was able to perform.

I quantitively analysed a total of 33 sets of scores from students' self-assessments, the mid-term exam results, the final exam results and their journal grades.

4.1.1 Descriptive statistics

The mean score for the self-assessments was 59.33, with a standard deviation of 10.7, whereas the mean score of the results for all assessments (midterm exam, final exam, and journals) rated by the teachers were 67.5, 67.5, and 72.3, respectively.

Table 2 Descriptive statistics of the data

	N	Minimum	Maximum	Mean	Std. Deviation
Self-assessment score/100	33	30.0	75.0	59.333	10.7752
Midterm CI exam score/100	33	50.0	80.0	67.500	6.6849
Final CI exam score/100	33	40.0	90.0	67.521	9.6653
Journal score/100	33	62.0	85.0	72.394	6.5141
Valid N (listwise)	33				

A comparison of the students self-assessment mean scores and the teacher's scores showed that 26 of the 33 students underestimated their performances in comparison with teacher-rated scores by an average of 11.9. Only four students overestimated (average 7.8 points), and two students accurately estimated their scores (both of them were in IP2Kr).

Table 3 Overview of self-rated scores compared with teacher-rated scores

	Underestimation			Overestimation		
	Number of students	Range	Average	Number of students	Range	Average
IP3Ch (*n*=9)	9	-1 to -25	-10	0	N/A	N/A
IP3Kr (*n*=11)	10	- 0.5 to -18	-9	1	N/A	3
IP2Kr (*n*=13)	7	-1 to -45	-18	4	2 to 14	9

4.1.2 Correlation between self-rating and teacher-rating on the midterm exam

Preliminary analyses, assessed by Shapiro-Wilk's test, showed a non-linear relationship between self-rating and teacher-rating ($p < 0.05$). Two outliers were detected where two students grossly underestimated their scores (IP2Kr3 by -45 and IP2Kr10 by -31). After removing these outliers, the scores were normally distributed. Nonetheless, the analysis was run with and without outliers to see if there was a difference in the results.

Table 4 Correlation (outliers included)

		Self-assessment score/100	Midterm CI exam score/100
Self-assessment score/100	Pearson Correlation	1	.138
	Sig. (2-tailed)		.443
	N	33	33
Midterm CI exam score/100	Pearson Correlation	.138	1
	Sig. (2-tailed)	.443	
	N	33	33

The results showed that the two scores, self-assessment score and midterm CI exam score, were not correlated as the *p*-value was greater than 0.05. The analysis run without the outliers also had a *p*-value greater than 0.05, suggesting there was no significant relationship between the two sets of the scores. The results indicated that the students' perceived scores of their performances were not similar to the scores given by their teachers.

In addition, the range of underestimation and overestimation was wider in the IP2 class than in IP3, as shown in Table 3. As mentioned earlier, two students in the IP2 class were outliers who underestimated their performance by 45 and 31 points respectively when compared with the scores given by their teachers on the same performance. The two students who overestimated their performance the most in this study were also in the IP2 class, with eleven and fourteen points respectively. This might indicate that experience in receiving marking from teachers gives students some sense of an appropriate range of scores as the students in the IP3 class were more accurate in evaluating their performance.

4.1.3 The relationship between the accuracy of self-assessment and performance outcomes

A multiple regression was run to predict final exam scores from the difference between self-ratings and teacher-ratings ("Difference"), journal scores, and the mid-term exam scores. The multiple regression model statistically significantly predicted final exam scores, $F(3, 29) = 34.267$, $p < .0005$, adj. $R^2 = .75$. Table 5 ANOVA, below, shows the statistical significance of the overall model.

Table 5 ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2331.605	3	777.202	34.267	.000[b]
	Residual	657.750	29	22.681		
	Total	2989.355	32			

A weak relationship between the variable 'Difference' and the final exam scores was also shown (-0.089). The results also show that the relationship was negative, as shown in Table 5 below.

Table 6 Summary of Multiple regression analysis

Variables	*B*	*SE B*	β
Difference	-0.073	-0.089	-0.89
Journal score/100	0.020	0.013	0.013
Mid-term score/100	1.214	0.839	0.839*

Note. * $p < 0.05$; *B*= unstandardized regression coefficient, *SE B* = Standard error of the coefficient; β= standardised coefficient

These results will be further explored and interpreted in Chapter 5, and discussed in the light of the findings of the qualitative analyses.

4.2 The results of qualitative analysis: Journal entries

The results are presented as charts based on the number of references for each parent or child node (a collection of references from the journals). The discussion of each makes reference to the qualitative analysis of the journals and the requirements of the assessment tasks. As the

comments made by the students to evaluate their performance were assigned to a matching criterion in the rubric for self-assessment, their use of the rubric was shown by the frequency with which the student discussed the criterion in their journal.

It should be noted that the number of references about self-assessment in the journals varies across students and classes. There are two potential reasons for this. First of all, students used different styles of writing because they were allowed flexibility in the format of their journal. Secondly, while the students in the IP3 classes received instructions for the assignment to self-assess, responding to all the criteria in the rubric, analysing causes, and seeking solutions for two major issues, the students in IP2Kr were guided to write a "brief analysis report" by selecting two criteria in the rubric. These differences in the instructions resulted in variations in the total number of references: IP3Kr accounted for 50% of the total number of references, IP3Ch, 34%, and IP2Kr 16%. This difference required careful consideration in analysis and presentation because the trend of the class with the largest number of references might outweigh the trends of the other two classes. In addition, it makes it difficult to compare the differences among the classes. Therefore, I will use percentages for each class rather than the actual number of references.

Students discuss most …

In this section, I will present the findings at the level of the two broad categories of the rubric:

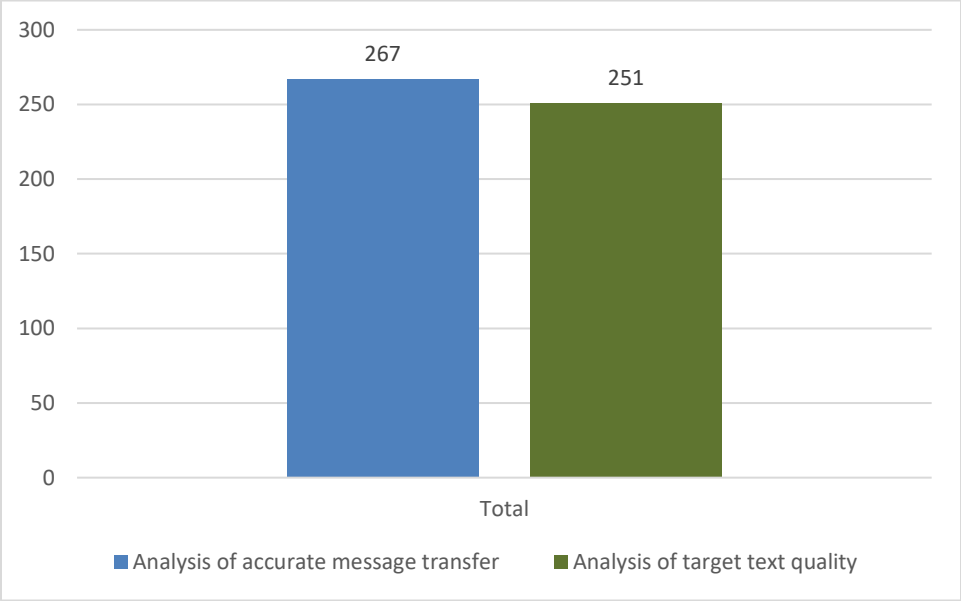
- "Analysis of accurate message transfer," and
- "Analysis of Target text quality."

I will firstly provide an overview and then discuss each criterion and sub-category of the criterion to document how students used the rubric in more detail. I provide the results for the

whole cohort to give a general understanding of the students in the program in. its entirety, and then provide results for each separate class (language pair and level) to see the differences and similarities between the classes.

4.2.1 Overview

Figure 7 Comparison of frequency between "Analysis of accurate message transfer" and "Analysis of target text quality"



Overall, the number of references assigned to "Analysis of accurate message transfer" and "Analysis of target text quality" were 267 (52%) and 251 (48%), respectively. There was no major difference in the number of references for each of the two categories. This suggests that students, on the whole, addressed both categories of the rubric in a balanced way.

When analysed by class, while the trend in IP3Ch and IP3Kr was similar to the overall trend in that there were slightly more references in relation to "Analysis of accurate message transfer" than "Analysis of target text quality," the trend in IP2Kr was the opposite. The share of 'Analysis of accurate message transfer' of the total reference in IP3Ch was 57% and 53% in IP3Kr, but only 36% in IP2Kr. This shows that the students in the advanced interpreting classes were more concerned about the accuracy of their interpreting than the target text quality, while

the students in the lower level of interpreting practice class were more concerned with their target text quality when they self-assessed their performances.

Figure 8 Overview by class

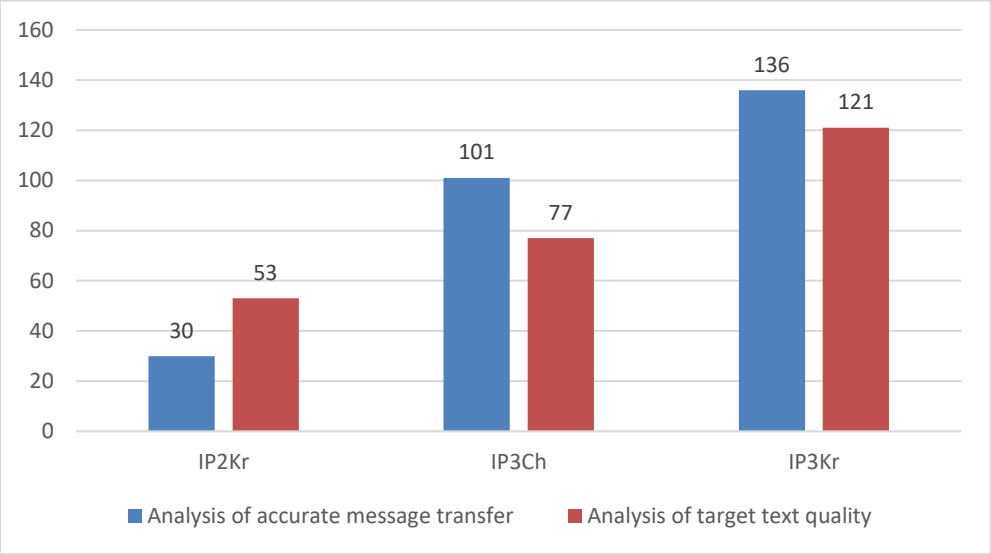


Table 7 The percentages of "Analysis of accurate message transfer" and "Analysis of target text quality"

	Analysis of accurate message transfer	Analysis of target text quality
IP3Ch (*n*=9)	57%	43%
IP3Kr (*n*=11)	53%	47%
IP2Kr (*n*=13)	36%	64%

The subsequent charts show the frequencies of the references in relation to performance for each of the sub-categories of the rubric. The sub-categories for "Analysis of accurate message transfer" are "omissions/additions/substitutions," "lexical/conceptual matches," "equivalent register," "equivalent effect," and "unspecified." The sub-categories for "Analysis of target text quality" are "fluency & naturalness," "grammatically correct," "comfortable pace/speed," "clear articulation/pronunciation," "coherence of TT message," "sufficient volume," and "unspecified."

4.2.2 "Analysis of accurate message transfer"

Figure 9 Distribution of references for sub-criteria under "Accurate message transfer"

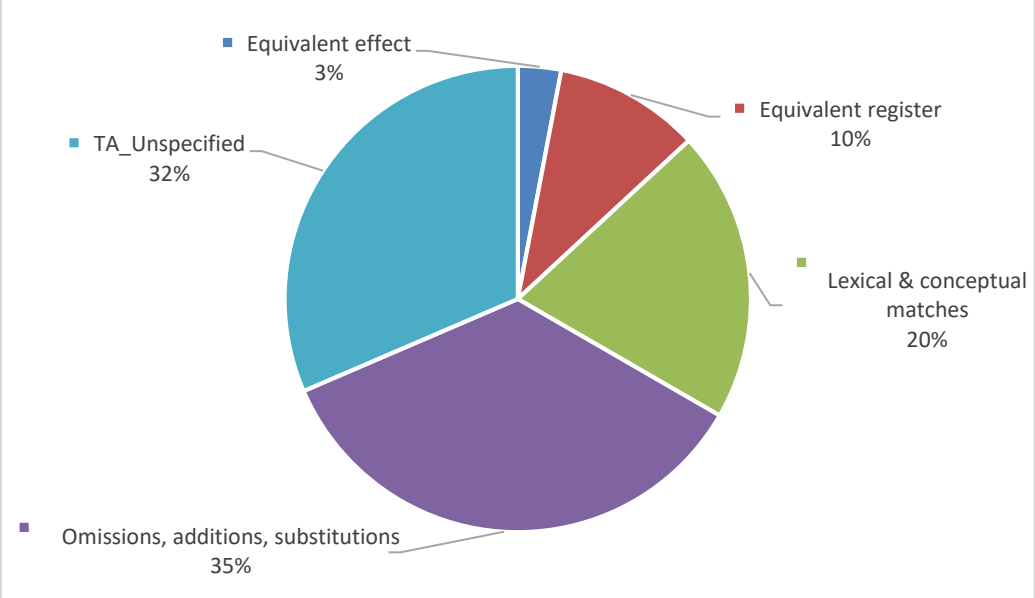


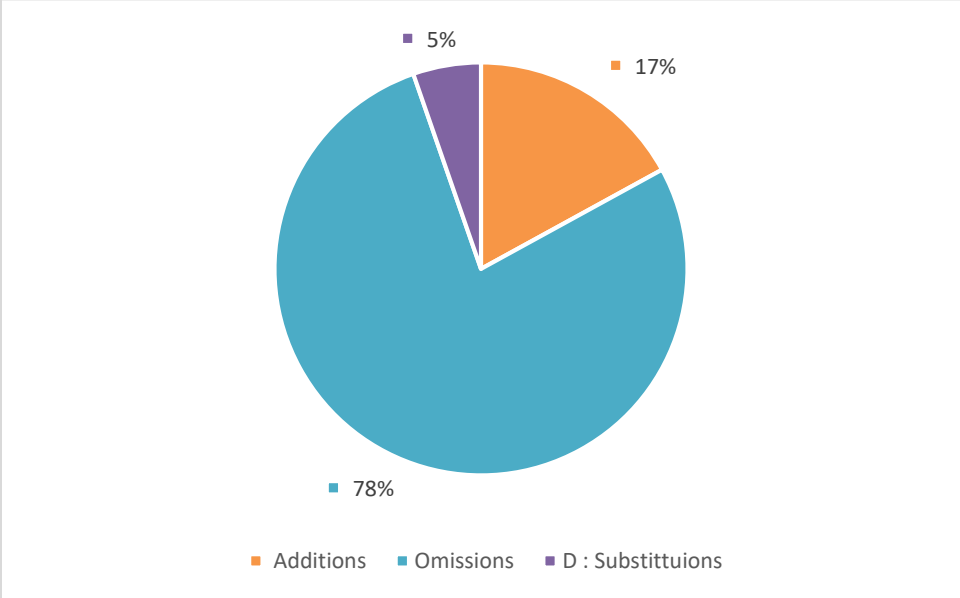
Table 8 Share of references for sub-criteria under "Accurate message transfer" by class

	Omissions/additions/substitutions	Lexical & conceptual matches	Equivalent register	Equivalent effect	Unspecified
IP3Ch (*n*=9)	39%	19%	8%	3%	33%
IP3Kr (*n*=11)	33%	18%	12%	4%	32%
IP2Kr (*n*=13)	33%	33%	10%	0%	23%

"Omissions/additions/substitutions"

Among criteria under "Analysis of accurate message transfer," the criterion "omissions/additions/substitutions" was the most frequently referenced at 35%. Looking at each of the classes it ranked first for all of the cohorts. This suggests that the students used this criterion the most in the evaluation of the accuracy of their performance. Since this criterion is a combination of three different types of error, the number of references for each error type was further explored.

Figure 10 Share of "omissions/additions/substitutions"



As shown in Figure 10, of the three error types in "omissions/additions/substitutions," "omissions" dominated the references. It accounted for 78% of the total references assigned to this node, while references for "additions" accounted for only 17% and "substitutions" only 5%. This trend was evident across all classes, except for IP2Kr, where "omissions" accounted for 100% of the references in this category (*Table 9*).

Table 9 Reference share in each class

Class	Omissions	Additions	Substitutions
IP3Ch (*n*=9)	74%	21%	5%
IP3Kr (*n*=11)	76%	18%	7%
IP2Kr (*n*=13)	100%	0%	0%

The term "comprehension" was used in the rubric. Based on the qualitative analysis of the journals, it is apparent that the students most often associated their problems of comprehension with omissions in their interpreting, citing it as the cause of the omission:

> "I omitted numerous sentences in the interpreting in that I was not able to comprehend the source text (English). For example, I missed the part of 'It's really a failure to detect a certain kind of danger'." (IP3Ch3)

Therefore, when students discussed both “Comprehension” and “Omissions/additions/substitution” in their journals, the number of references to omissions had to increase as it appeared in relation to both sub-criteria.

In terms of “additions” or “substitutions,” these error types usually occur as the result of a strategic choice. It is likely that the students were not able to employ a strategy when they encountered a problem during their performances. Indeed, references about “additions” were in the journals of the students whose mid-term mark was above the average. More competent students seemed to be able to use the strategy, or were aware of using it. The reference below comes from the journal of the student who received the highest mark at the mid-term exam.

> “In case my target audience might not get the point, I even added ‘这些只是开个玩笑，但是 (These are only jokes, but) ……’ to help the audience understand. I found myself prone to make additions when I was not able to express the meaning of an expression.” (IP3Ch6)

Another possibility is that the students thought their “additions” and “substitutions” were acceptable or successful, and thus did not feel the need to discuss the criteria in their journal.

“Unspecified”

References assigned to the node “unspecified” accounted for 32% of the total number of the references under “Analysis of accurate message transfer,” the second largest share in the category. As explained in the coding process in section 4.1, references that include “misinterpreting,” “distortion,” or generally discuss the accuracy without referring to other specific criteria in the accuracy category were assigned to this child node. It shows that the students evaluated the accuracy of their performance using the words “misinterpreting” or

"distortions," although the words were not in the rubric. For example, some students reported cases where they converted negative messages into positive messages:

> "I made some major mistakes in the second part such as mistranslating "a super intelligent AI that was no smarter than your average team of researchers at Stanford or MIT" as "an AI that is much smarter than its researchers" (IP3Ch1)

In addition, similar to the references that are related to "omissions," the unspecified references were often associated with "comprehension." Therefore, when they addressed the criterion "comprehension," this type of reference was found in their journal.

> "Furthermore, I could not understand the key word of this part is the spectrum. Therefore, I just listed fragments of information without understanding the whole context, so the message of the TT is not accuracy at all." (IP3Kr14)

"Lexical/conceptual matches"

This criterion accounted for 20% of the references under "Analysis of accurate message transfer." Evaluating their performance in this aspect, the students provided examples where they failed to deliver an equivalent utterance in the target language.

> "As for lexical matches, I also made many mistakes. For example, 'failure of intuition' was interpreted into '想要失败 (intend to make failures)'." (IP3Ch20)

However, apart from when using "lexical/conceptual matches" as a heading in their journals, the students did not use the term "conceptual matches" but did use the term "lexical matches" explicitly when making comments under the criterion as in the example above. This leads us to question whether they had the ability to identify an example of "conceptual matches" in their performance. Another possibility is that they did not see any instances of problems with 'conceptual matches' in their performances. However, even in the journals where some students

addressed each criterion and expressed their satisfaction with their performances, the term was not mentioned.

"Equivalent register"

This criterion accounted for 10% of the total references of "Analysis of accurate message transfer." In terms of the number of students, fifteen out of the 33 students made comments about this criterion when evaluating their performances. In the references, the students provided examples that they thought were not appropriate for the register or explained how they managed to match the register of the ST to the TT.

> "In addition, due to my limited English vocabulary, I do not use high register words or expressions for formal interpreting. In this mid-term interpreting, I said ~whatever, and ~ something like that. These expressions are not appropriate to be used in formal text." (IP3Kr21)
>
> "Generally speaking, the TT was formed by formal sentences and delivered in a calming tone, thus it did create an equivalent affect and register as the ST." (IP3Ch2)

As the above excerpt from IP3Ch2 shows, there were relatively more positive references found for this criterion than for the others. It was found that "equivalent register" was the criterion that had the highest number of positive references (8 out of the 27 references, 30%) for the criterion of "Analysis of accurate message transfer." These references come from the journals of those students who made more positive comments, relative to other students. In other words, while most students wrote about their mistakes in the journal, some students included positive aspects of their performances, and "equivalent register" was one of the criteria they thought successfully performed. Therefore, it might be possible that other students who thought they managed to match the register of the ST to TT did not write about it and focussed on their mistakes in their journals.

"Equivalent effect"

This criterion accounted for 3% of the total references under "Analysis of accurate message transfer." Only five students discussed this criterion. It was noticed that some justifications the students provided for their judgements in regard to this criterion were not clear or sufficient to determine whether they have evaluated this aspect with a clear understanding. The examples of IP2Kr4 and IP3Kr14 show that they related mistranslation to "equivalent effect." To some extent, mistranslation could negatively affect "equivalent effect," but it seems that they made a judgement based on one mistake, while the criterion needs a pragmatic approach to analyse.

> "접속사의 잘못된 선택으로 오역이 발생했습니다. (I mistranslated this part by using a wrong connective.)" (IP2Kr4)

> "I could not hear 'if I were to…' and 'wouldn't' and it resulted in a critical mistake. Therefore, the equivalent affect is bad." (IP3Kr14)

In the following section, I present what students discussed under the second category, "Analysis of target text quality" and each of the sub-categories within this criterion.

4.2.3 "Analysis of target text quality"

Figure 11 Distribution of references for sub-criteria under "Target text quality"

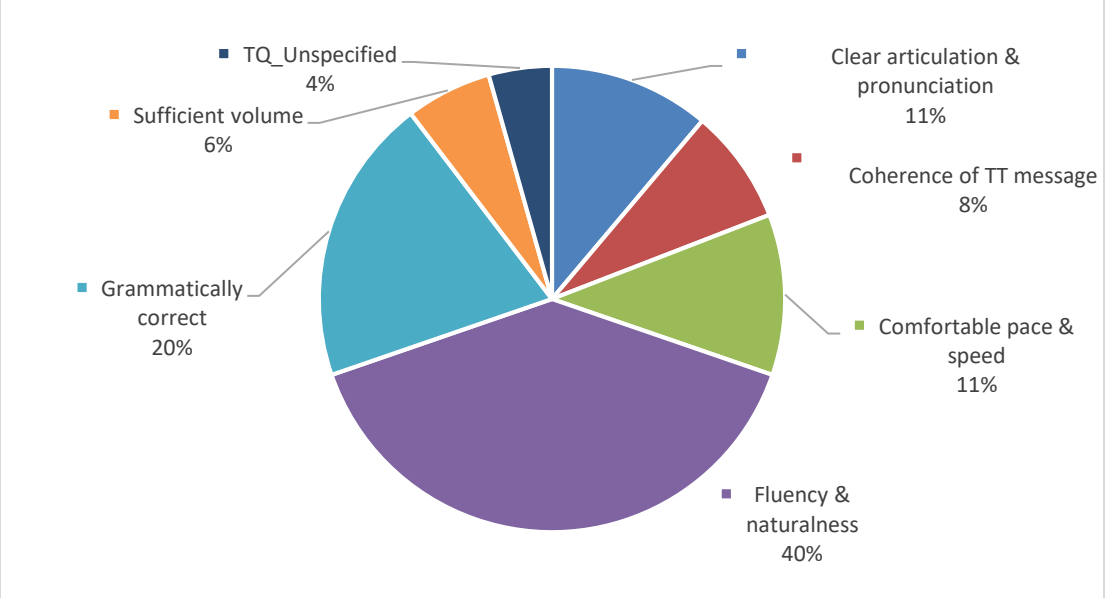


Table 10 Share of references for sub-criteria under "Target text quality" by class

Class	Fluency & naturalness	Grammatically correct	Comfortable pace & speed	Clear articulation& pronunciation	Coherence of TT message	Sufficient volume	Unspecified
IP3Ch (*n*=9)	35%	16%	10%	12%	12%	6%	9%
IP3Kr (*n*=11)	38%	19%	11%	13%	9%	8%	2%
IP2Kr (*n*=13)	49%	28%	13%	6%	0%	0%	4%

For "Analysis of target text quality," the most frequently discussed criterion in all the classes combined, as well as each class, was "fluency and naturalness." "Grammatically correct" followed next, and the least discussed was "sufficient volume." The "unspecified" included general comments about target text quality without referring to a specific criterion or an explanation about the judgement of the quality. For example,

"The TT quality in this task was relatively better." (IP3Ch10)

Such references were found when students wrote a general comment about the target text quality and it was usually followed by references that addressed a specific criterion in the target text quality category.

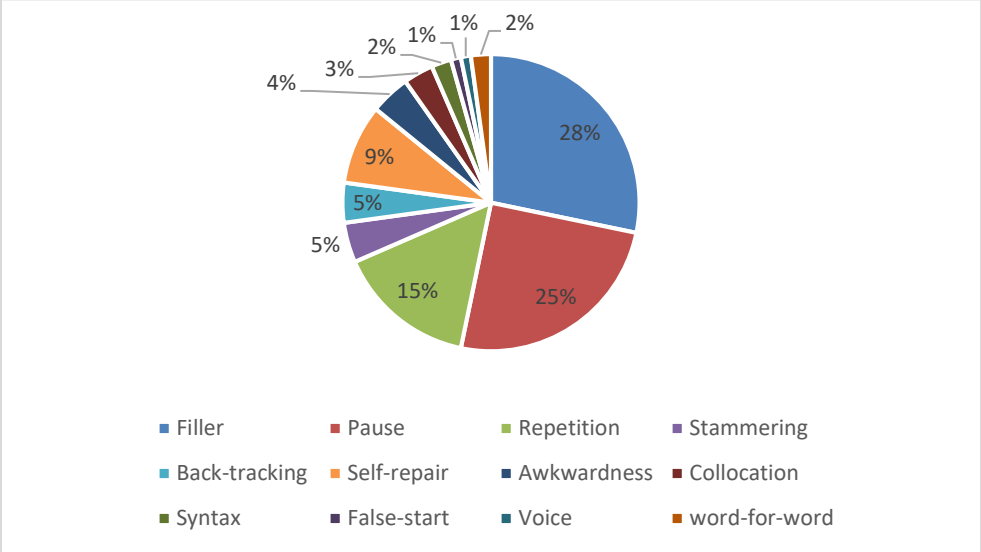
"Fluency & naturalness"

The references related to "fluency & naturalness" accounted for 40% of the total references. In each class separately, it accounted from 35% to 49% of the total references. When evaluating their performance in terms of "fluency & naturalness," the students often provided descriptions (sometimes excerpts from transcription) and/or used specific terms to justify their judgement. For example,

> "However, the interpreting was not fluent due to many fillers such as 'um', 'so', 'ah', 'hmm'." (IP3Kr16)
>
> "false-start and self-repair (TT: 一個就是我們會 (firstly we are going to)…第一道大門會是 (The first gate would be)) occurred in the next utterance due to my attempt to recall the memory of the ST content through the notes." (IP3Ch10)

The students used a range of terms to evaluate this sub-criterion. Based on their descriptions and the terms used, those comments were sub-categorised under "fluency and naturalness." A total of twelve sub-categories from 92 references were made: fillers, pauses, repetitions, stammering, back-tracking, self-repair (or self-correction), awkwardness, collocation, syntax, false-starts, voice, and word-to-word translation. Among those, the sum of references for the three sub-categories accounted for 68%: fillers 28%, pauses 25%, and repetitions 15%.

Figure 12 Sub-categories under "fluency & naturalness"



In case one specific term might have been intensively used by only a few students, the number of students who discussed these terms was analysed.

Figure 13 Constructs of "fluency & naturalness" used by the students

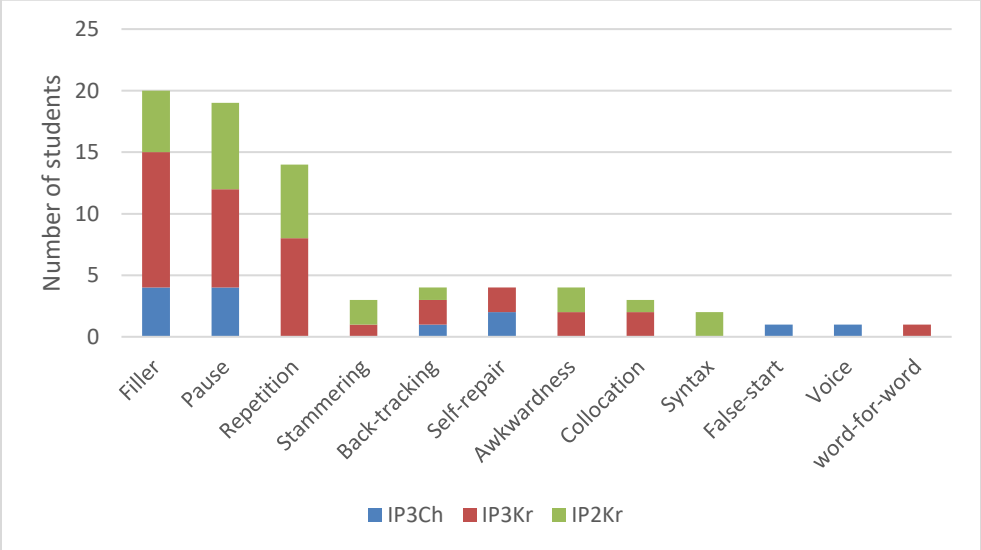


Figure 13 shows that while fillers, pauses, and repetitions appeared in the majority of the students' journals to discuss "fluency & naturalness," the rest of the sub-categories were mentioned by fewer than five students. This result indicates that the students focussed heavily on fillers, pauses and repetitions when evaluating their "fluency & naturalness." It might be that the

three were the most serious issues that they felt negatively affected the target text quality for this criterion. However, it is also possible that most of the students looked only for these constructs, which are easier to detect and/or report than others, such as collocation and word-for-word interpreting.

It was also noticed that some students distinguished "fluency" from "naturalness" although the two were combined as one criterion in the rubric. It is interesting that the students related a specific term to a different criterion. For example, seven out of 20 students discussed their filler issue distinguishing "fluency" from "naturalness." While six of them put the issue under "fluency," one student associated it with "naturalness." The excerpts below are from the same class. The examples show that students have different ideas about the constructs of "fluency" and "naturalness," even when they are from the same class.

> "The target text is fluent for the most part except some repetitions of "Er…." (IP3Ch20)

> "…hesitations markers of 'er' was inserted between the adjective and noun, which made the utterance unnaturally developed." (IP3Ch10)

Another case is that a student discussed his/her filler problem under the heading "omissions/additions/substitutions."

> "Ah⋯.the global competitors rise, 4th industrial revolution and ah…oil and […]. 문제점: 역시나 기호를 해석하느라고 중간에 아 아 어어 하는 소리를 많이 냄. (The problem: while decoding symbols (in my note), I frequently made sound of ah, ah, uh, uh.)" (IP2Kr14)

Similar disagreement on the constructs of the criterion was shown in the use of other terms as well. Among fourteen students who talked about "repetitions," two students addressed the issue problem of "naturalness" and one student related the issue to the "fluency" of the performance. Likewise, one student linked "back-tracking" to "naturalness" and another student discussed it

under “fluency.” When it comes to “pauses,” eight of the students discussed it under “fluency” or “fluency and naturalness,” while five students related the issue to “comfortable speed/pace” (and these comments were categorised as speed/pace). It seems that the students are able to identify and explain an issue using a specific term, but have different ideas about the constructs of “fluency,” “naturalness.”

“Grammatically correct”

This criterion came in second place in “Analysis of target text quality,” following “fluency & naturalness” with 20% of the total references for the category. Most of the references were about grammatical mistakes with specific examples. There were few positive references regarding this criterion. The expressions used in these references were “…with less grammatical mistakes,” “grammatically acceptable,” and “TT is generally grammatically correct.” It seems students have their own yardsticks when judging their performances in this aspect. How many errors there were, or what type of grammatical mistakes were acceptable for them, remains unclear.

“Comfortable pace/speed”

Eleven per cent of the total references regarding target text quality were concerned with this criterion. Similar to the case of “fluency and naturalness,” the students used either “pace” or “speed” when they addressed this criterion.

> “Speed is not properly managed. As the comprehension of ST was poor, it took time to read and understand the note. That leads to many pauses in a sentence.” (IP3Kr8)
> “However, I found my rendering pace was the biggest issue among all the problems I had. I paused too long between each word, […]” (IP3Ch1)

Although these two students use different terms, it seems that the two terms were used synonymously. It is interesting that the students selectively used one word to address the criterion. There were five students who used "rate" instead of "pace" and "speed."

"Clear articulation/pronunciation"

This criterion also accounted for 11% of the total references about target text quality. As above for "speed" and "pace," the students used either "articulation" or "pronunciation" in their journal. There was one student who used both and seemed to understand the difference between the two.

> "However, when I listened to my record, I found out my small and unforgivable pronunciation mistakes such as the pronunciation of the word 'historical [hɪ'stɒrɪk(ə)l]'. I had clear articulation and sufficient volume though, I could pace up a little bit." (IP3Ch3)

The above example leads to us to question whether the students used one of the two terms with an assumption (or knowledge) that the two terms are synonymously used in the case of "comfortable speed/pace" and "clear articulation/pronunciation."

"Coherence of TT message"

Coherence of TT message accounted for 8% of the total number of references. While it took up 12% of the references of the journal entries in IP3Ch and 9% in IP3Kr, there was no reference related to this criterion in the IP2Kr journal entries. This might be due to the instruction given to the students; they were required to provide two sentences as examples showing the aspects which the students felt were the most problematic in their performances. Therefore, it is likely that the students in the class focussed on issues they could find at word or sentence level, but not at context level, which is necessary to evaluate their output in terms of coherence.

The references about this criterion were associated with accuracy-related criteria, such as "omissions."

> "There are quite a lot of omissions in the target text as I could not follow the speaker's speed while taking note, therefore the coherence of target text was not achieved." (IP3Kr8)

Thirteen out of twenty students (IP3Kr and IP3Ch) made references to this criterion, though only three students made positive comments about it. Considering that "omissions" was a major issue among the students, it seems that most of the students who responded to "coherence" were not satisfied in terms of this criterion. I noticed that student IP3Kr8 evaluated his/her performance as successful in terms of "coherence of TT message," justifying that omissions did not affect "coherence" while the other two students did not provide justifications.

> "As the ST comprehension was alright, coherence of TT is achieved. There is no sudden change of topic by mistake. Though some details are missing, I believe that the message from the speaker is coherently mentioned in TT." (P3Kr8)
> "The message I transferred was coherent and clear." (IP3Ch6)

Among the seven students who did not make any comments about this criterion, five of them were those who did not make any positive comments in their journals. Therefore, it is possible that they did not respond to this criterion because they thought it was successfully accomplished in their performances. This cannot, however, be determined, since the other two students who made positive comments on other aspects of their performances seemed to skip the criterion.

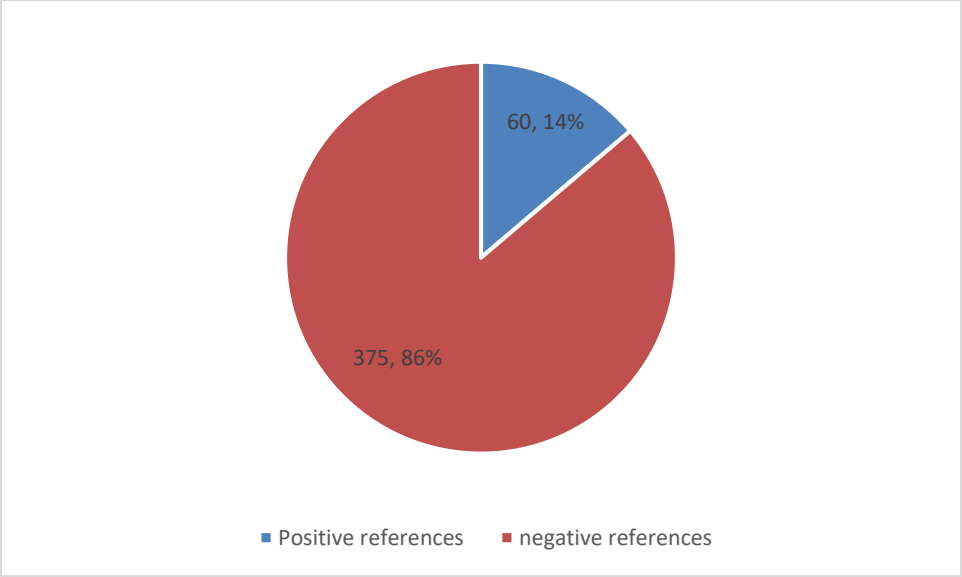
"Sufficient volume"

"Sufficient volume" accounted for 6% of the total references about "Analysis of target text quality." Unlike other criteria, most of the references about this aspect were positive (twelve out of fifteen references, 80%). For this criterion, there was no reference in the journals of the IP2Kr class. This might be the reason why this criterion had the least number of references. Another potential reason is that this is a positively worded criterion, and since negative

comments are more prevalent, those who did not discuss any positive aspects in their performances skipped this criterion.

4.2.4 Positive comments

Figure 14 The share of positive references



The students tended to make more comments about the negative aspects of their performances than the positive aspects in their journals. The students in IP2Kr were guided to discuss only their mistakes and analyse the causes. Therefore, there were no positive references in their journals and all the positive references were made by the students in the IP3 classes. Only 60 of the 435 references (14% of IP3Kr and IP3Ch) were positive comments. Among these positive references, 21 references were about "Analysis of accurate message transfer" and 39 references were about "Analysis of target text quality." For "Analysis of accurate message transfer," "equivalent register" had the highest number (seven references). For "Analysis of target text quality," "sufficient volume" had the highest number (twelve references). It was noticed that all the positive references were made by twelve of the 20 students. Eight students did not make any positive comments about their performance. Also, those who made positive

references seem to try to use most of the criteria in the rubric. The average number of criteria used by these students were nine (there were eleven sub-categories). Therefore, it may be interpreted that the eight students who did not make any positive comments might have excluded the criteria that they thought were successfully performed from their journal entries.

4.3 The results of qualitative analysis: Interviews with the lecturers

This section presents one-on-one semi-structured interviews with three teaching staff of the classes in this study. From the interviews, I was able to hear how they communicated with the students when giving the assignment, as well as their thoughts about and experiences of the assignment. The teachers confirmed that the journal-writing assignment after the mid-term exam was the only assignment related to self-assessment in the semester. Even so, they said that they encourage students to self-assess for their out-of-class practice.

Q1. Apart from the assignment, do you guide students about self-assessment for their self-study? If so, what guide do you give them?

Amber[5] provides an article about student self-assessment ability on iLearn[6] to her students and asks them to use the information. The article includes a rubric that is different from what is used for the assignment. Nevertheless, it was a recommendation, rather than mandatory. Similarly, the other teachers do not make self-assessment a mandatory activity outside of this particular assignment. Jessica advises the students to record and listen to their recordings. Claire gives advice to her students to self-assess as well as peer-assess. Claire explained that she had encouraged self-assessment but noticed her students were reluctant to practice it. She added that

[5] Pseudonyms are used for the staff to ensure anonymity

[6] iLearn (http://ilearn.mq.edu.au) is Macquarie University's online Learning Management System (LMS). It provides an online environment for learning, teaching, communication and collaboration.

her students wanted her to assess their performance and did not understand why they had to self-assess. That is why she also suggests peer-assessment to her students. Since doing self-assessment was not compulsory, it was up to the students to seek further feedback or advice about their self-assessment, and the teachers did not have to check whether the students practiced self-assessment. Therefore, the assignment was the only opportunity for teachers to see how students practiced self-assessment.

Q2. What instructions and rubric do you provide to the students for the assignment?

The answers from the teachers were similar to each other in that the way they provide the instructions and rubric. According to their answers, it was clear that their focus for the assignment was on students' practice in cause-analysis and strategic planning for improvement, rather than self-assessment.

> "[…] the second task is writing a self-assessment report. No. It is called reflective journal. So, it has a self-assessment component, but it is more like self-reflection. So, students have to write a report on the basis of how they went at the mid-term exam and they have to identify the causes behind the mistakes that happened during the exam and why they happened and how they think they can prevent the same things from happening in the future." (Amber)

Jessica explicitly focussed on students' cause-analysis skills and strategic planning. Her instruction suggests that she is aware of the importance of modelling in enhancing students' understanding.

> "I stressed that it was about analysing. For example, I found that students usually said, 'I did this wrong,' 'I omitted this,' and 'I will put more effort.'" Therefore, I gave the instruction with examples so that they identify concrete causes and strategies for improvement." (Jessica)

Claire did not use the term "self-assessment" when she explained the assignment for this question in the interview, which suggests that her focus was not on self-assessment practice.

> "It is all in the unit guide. So, when we meet for the first time everybody would be given a copy of unit guide which is also published on iLearn. […] I make some brief explanation in class if you have no clue as to how to write a reflective journal at all. […] I guess they should have a good idea as to how a journal should be looking like." (Claire)

Q3. How do you explain how to use the rubric?

In terms of the rubric, all of the teachers answered that they provided the rubric to their students in advance but did not feel the need to explain each of the criteria in detail in the rubric. They also said the student would have asked if they did not understand the rubric, and they added that they had not received any questions from their students about the criteria.

> "No, I don't. It is quite self-evident. If they have any questions, they would ask, right? I haven't received any questions so far. So, I guess that means they are very clear." (Amber)
>
> "No, I didn't explain it. I handed the rubric out and told them to write about one criterion from each of the two areas. There was no student asking about the criteria later. I think they did alright." (Jessica)

Claire had the same idea as other teachers that students should have a good understanding of the criteria, and the criteria are "self-explanatory." When answering this question, she went through the rubric placed on the desk. When she saw "conceptual matches," which was not used by any of the students in their journals, she seemed to become aware of the possibility that the students might not be familiar with the criterion.

> "I would say all of them have a good idea what these terms mean. So, I don't necessarily have to explain these terms in detail. But I offer if you got any questions, feel free to ask and I would explain that. […] These are self-explanatory. You don't have different opinion in terms of comprehension and the conceptual matches… umm this [pointing at the word "conceptual matches" in the rubric] might be something they would ask." (Claire)

This assumption by the teachers led to the next question.

Q4. When giving in-class feedback for your students' performance, do you use the terminology in the rubric?

Amber and Jessica said they think they use the terms in their class while Claire said she does not necessarily refer to a criterion when giving feedback in her class. The answers give an impression that the teachers were not necessarily conscious about their use of the terms for quality assessment and the alignment of the rubric when giving feedback to their students in class.

> "I guess, so. I have to. Although I don't consciously do that …umm yes…" (Amber)
> "I cannot say I use the exact wording of every criterion, but I guess I use most of the wordings. For example, grammar, pronunciation, voice, omission, misinterpreting… yes… speed…." (Jessica)
> "Normally I would give them feedback based on what I've heard to be problematic. I don't necessarily use the terms in the rubric, but I probably would point out exactly what went the problem is." (Claire)

It can be inferred from Claire's answer that the students might have the ability to evaluate the aspects that they did not use in the journal because they might have learned from their teacher's in-class feedback. However, if students were not familiar with the wording of the criterion, it is possible they could not use a criterion in the journal, and thus chose not to discuss that particular aspect.

The next questions were about the teachers' feedback on the assignment and the interpreting performance.

Q5. In relation to the assignment, what does your feedback include?

There was a difference in the way the teachers give feedback. Amber provided written feedback on performance, reflection, and academic writing skills. She said her feedback focusses more on the report (reflection) than exam performance, and lets students know what she thinks

their strengths and weaknesses are. Her feedback on performance is holistic. Jessica said that she provides written feedback on both performances and reflections. She points out mistakes in a corrected version that students provide in their reports if they are incorrect. Both Amber and Jessica said they also suggest solutions to students' problems. Claire said that she provides written feedback on reflections and academic writing skills. She added that she cannot give feedback on students' self-assessment because she cannot know how her students self-assessed by reading their journals. She also noted that providing feedback on self-assessment requires one-on-one consultation.

Q6. Do you release the interpreting mark to the students?

Releasing the mark was up to the teachers in the Interpreting 3 classes. Teachers in the classes let the students know upon their request.

> "I don't. I just… I don't know… I thought it would be better for them. I thought that just feedback itself is enough because there's always mixed responses. Some students want to know, but some students don't. So, I let things be democratic. If they really want to know, they can always request it. I am happy to let them know. But if they don't want to, then, no. They would be silent. I let them be." (Amber)

In the IP2 class, the teacher explained that it is the policy of the unit not to release the mark to the students.

4.4 Summary of the results

In this chapter, I presented the results of the data analysis (ratings, journals, and the interview). The analysis of the ratings showed that the students' self-rated scores were not similar to those of their teachers and had a weak and inverse relationship with their final exam performances. The analysis of the journals indicates that, overall, students used most criteria in the rubric to evaluate their performances. However, they gave more focus to certain criteria, such as "omissions/additions/substitutions" and "fluency & naturalness," while "equivalent effect"

and “conceptual matches” were hardly mentioned, if at all. Also, it was also found that the majority of their comments were about the negative aspects of their performances.

The analysis of the interviews with teaching staff showed that their ideas about the assignment and their instructions were quite similar. They focussed on students’ cause-analysis and future improvement plan rather than students’ self-assessment practices, with an assumption that the students have a good understanding of the criteria in the rubric and are therefore able to self-assess without much guidance.

CHAPTER 5: DISCUSSION

This study set out to explore the student's ability to self-assess, the relationship between the accuracy of students' self-assessment and the results of their performances in the final exam, and lastly, the relationship between students' ability in self-assessment and the teacher's instruction. Chapter 2 reviewed the literature on general education and interpreter education in regard to the role of self-assessment in students' learning and skill development. Chapter 3 examined the methodological approach and described the methods for data collection and analysis employed in this study. Chapter 4 presented the results of the data analysis. This chapter examines and interprets the results of this study. Finally, I outline some limitations of this study, and make suggestions for further research.

5.1 Students' perceived performance quality

5.1.1 Underestimation and focussing on negative aspects

This study found that the majority of the students underestimated their scores. This tendency was pervasive in the advanced interpreting classes. This could be explained by the finding from the qualitative analysis of the journals that 86% of the journal entries were negative comments about the students' performances, and the students tended to provide detailed information with examples of their mistakes. Bartłomiejczyk's (2007) study shows that interpreting students tend to focus more on negative aspects than positive aspects even when guided to focus on both negative and positive aspects. The students in this study, however, were instructed to self-assess their performance and analyse mistakes. The guidance might have reinforced the tendency because it seems that there was no clear distinction drawn between self-assessment and cause-analysis in the instructions. As Andrade and Valtcheva (2009) noted, the primary purpose of self-assessment is to identify a learner's strengths and weaknesses to promote

learning. The students in this study did not identify their strengths explicitly but focussed on mistakes in their journals. In this sense, it could be said that they were not able to make good use of self-assessment. If they not only did not write about their strengths in their journals, but also did not try to identify them in the process of their self-assessment, it may be the case that their self-assessment practices should be reviewed. This is because if students perceive they are constantly doing badly, they are more likely to be demotivated and thus find it hard to sustain their efforts (Z. Wu, 2016), which in turn adversely affects their self-regulated learning (Zimmerman & Schunk, 2011). A positive focus not only leads to improvement in learning, but acknowledging strengths has a positive psychological impact that helps interpreting students ease their anxiety (Kim & Chiu, 2011).

5.1.2 Self-assessment and the outcomes of teacher assessment

Although the results of the multiple regression analysis showed a weak relationship between self-assessment and the outcomes of the teacher assessment, the inverse relationship would seem to indicate that students who are harsher in scoring themselves are more likely to obtain a higher score in the final exam, particularly in the cases of those who grossly underestimated their results. This result was unexpected and rejects the hypothesis that more accurate self-assessments correlate with improvement in the outcomes, as the literature in self-assessment has suggested (Andrade et al., 2008; Ross et al., 2002).

For example, one student who gave him/herself 45 points less than his/her actual score (the average underestimation was -11 points) improved by eight points in the final exam (the average improvement was four points) compared to the mid-term result. Considering that students seem not to have received structured guidance in self-assessment, it is possible that this student evaluated his/her performance against a higher standard than other students or a personal

ultimate goal, which would inevitably result in gross underestimation. Interestingly, it seems that the student used the disappointment as a motivation to put more effort into improvement. This appears to indicate that there might be psychological factors affecting interpreting students' self-assessment behaviours and their impact on the outcomes, which lies beyond the scope of this study.

5.2 The students' application of the rubric

The analysis of the journal entries showed that, in general, the students evaluated and discussed their performance by referring to most of the criteria and providing one or more concrete examples. It was shown that they have the knowledge of specific constructs of a criterion. For example, they discussed "fluency" relating to its constructs such as "fillers" and "pauses," although these specific terms were not provided in the rubric. This phenomenon is in line with what Hartley et al. (2003) observed in their study that interpreting students gain notions about interpreting quality and the ability to describe interpreting performance as they progress through their training.

However, students' use of criteria was heavily focussed on a few criteria, such as "omissions," "lexical matches" and "fluency & naturalness," as reported in Chapter 4. This indicates that the students were more concerned about these criteria than others when evaluating the quality of their performances. References about "omissions" and "lexical matches" stayed at word level and, in general, the students seemed to report their mistakes without discussing the impact of them on the overall quality of their performances. In addition, it might suggest that the students thought that rendering equivalent words is deemed successful, and if they could not, they perceived it as a mistake. If this is the case, such a mechanical perspective (Blenkinsopp & Pajouh, 2010) shown as the interpreting students' perception of successful interpreting

performance should also be reviewed, since omissions, additions, and substitutions could be used strategically for effective communication (Napier, 2004; Slatyer & Chesher, 2007).

In terms of "fluency & naturalness," the constructs discussed in this context of this criterion were limited to "fillers," "pauses," and "repetitions" in the majority of the journals. If the students could associate the criterion only with these constructs, it is likely that they overlooked other constructs related to it, such as self-repair and collocation, which would undermine their ability to self-assess their performance. Considering that quality constructs have not yet been clearly established in the minds of the students, as Fernández (2013) points out, students might not have been exposed to enough metalanguage to describe the quality of "fluency & naturalness," and thus focussed on concrete constructs as they were required to respond to the criterion in their journals.

5.2.1 Selective application of criteria

It was observed that the students discussed their performances selectively in their journals using the criteria that had been provided in the rubric, although the instruction for the journal writing required the students to "make sure you respond to the assessment criteria below." This behaviour by students needs to be further examined to determine whether it is related to the students' ability to apply the criteria or some other factor. There could be numerous possible reasons for the students' choices about which criteria to discuss in their journals. Also, a student might have a reason for not using one particular criterion and a different reason for another. Although the researcher could not follow up with the participants for further explanation of the results due to the very limited duration of the Masters of Research program, we can still contemplate the reasons based on the data and the literature. Here, I attempt to discuss the most likely reasons behind the students' selective use of the criteria for their journal writing

assignments. I lay out three broad categories first and then discuss each in more detail. The first probable reason is that the students did not clearly understand the assignment instructions. The second probable reason is that they did not understand the value (or concept) of self-assessment. The third probable reason is that they were not able to self-assess according to the aspects of those criteria that were not addressed in the journals.

Firstly, it is possible that they did not clearly understand the assignment instructions that they had to discuss their performance in relation to all the criteria. As the teachers' focus was on the students' cause-analysis and solutions, as shown in the interview, those students who did not discuss positive aspects at all might have thought that they did not need to discuss those aspects of their performances. However, this speculation cannot be applied to those who mentioned positive aspects, but still left out some criteria. Another possibility is a misunderstanding of the instructions. They might have understood that the assignment is only about self-analysis, and thus focussed on certain criteria that they thought represented the more serious issues in their performances. However, this explanation does not seem to apply to those who made positive comments but still left out only a few criteria while responding to most of the other criteria.

The second possibility – that the students did not understand that the value (concept) of self-assessment – can be viewed from two different perspectives. One is that they did not know the purpose of self-assessment is to identify not only weaknesses but also strengths. This is another probable reason for the decision of those who did not provide any positive comments about their performances. The difference between a misunderstanding of the instructions and this explanation is whether the students did find positive aspects but only selectively reported, or whether they only tried to identify weaknesses. The other perspective is that they might have perceived self-assessment as a "meaningless task" of no benefit, and became strategic by putting

in minimal effort (Maguire et al., 2001). In this case, students might have avoided a criterion that requires deep analysis of the source and the target text, and thus more effort. For example, "omissions" and "fillers," which all the students discussed, are problems that can be easily detected, whereas "equivalent effect," which only five students discussed, requires a pragmatic level of analysis (Pöchhacker, 2002).

The third possibility concerns the students' abilities in self-assessment. As Olina and Sullivan (2004) note, students might not have had enough confidence in their judgement using some of the criteria. If students are not sure about their judgements, they would choose not to discuss their performances in terms of the criterion because their assessment skills are to be evaluated as a part of the assignment fulfilment, which might be a negative backwash effect (Biggs & Tang, 2011). In the same vein, it is also possible that the students could not apply a criterion for self-assessment due to a lack of understanding of it.

For example, the criterion "conceptual matches" was not used by any of the students in the journals. The case may be that the students were either not confident about their judgement or unable to use the criterion. I make this suggestion because some students responded to all the criteria in the rubric except "conceptual matches." Also, potentially, the students might not have realised that they had skipped the criterion since it was put together with "lexical matches" in the rubric as "lexical/conceptual matches," and could have been missed if they did not have a clear understanding of "conceptual matches." "Equivalent effect" is a similar case to "conceptual matches" in that only five out of the 33 students discussed it. These two cases of "conceptual matches" and "equivalent effect" might be an implicit indication that some students might not have responded to some criteria because they were not able to assess their performances on those aspects.

5.3 Teachers' instructions in self-assessment

The literature that was reviewed in this study highlights the importance of a teachers' role in students' self-assessment skill development, and provides the conditions for this skill development from a pedagogical perspective. I learned from the interviews with the teachers in this study that their instruction in self-assessment was not ideal to meet the conditions that self-assessment theorists recommend for successful students' self-assessment (Andrade & Valtcheva, 2009; Panadero & Alonso-Tapia, 2013; Ross, 2006). I will discuss below the findings about the teachers' instruction in self-assessment in comparison to the conditions in the literature.

Prior to the interpreting performances and self-assessment task, the teachers provided the students with the criteria for self-assessment to guide them in the aspects of their interpreting performances that they needed to evaluate, which to some extent fulfils the conditions described in the literature. However, they did not explain the rubric because they thought the wording was clear and did not receive any questions from their students, which led them to assume that students were able to use the criteria without further support. Andrade and Valtcheva (2009) note that we cannot expect students to know and use assessment criteria when they simply receive a rubric, and thus teaching how to apply them is critical (Gibbs, 2010). The teachers probably arrive at this assumption because some of the words are indeed self-explanatory (e.g. grammar, omissions, pronunciation) and/or are used in class when they give feedback about their students' in-class performances, as the interviews with the teachers revealed. However, since they did not aim to link their use of the criterion in their feedback to the instructions for self-assessment, it remains unclear whether all the criteria were addressed in their feedback with examples which would have provided clear direction for the students when they needed to apply them themselves. Also, they might have described a mistake in general language rather than referring

to the criterion in the rubric, as Claire explained in her interview. This approach, however, may be problematic for students when they need to make a judgement in relation to specific criteria that are quite abstract, such as “equivalent effect” and “conceptual matches,” until they become capable of using them confidently. These are concepts that “do not have sharp boundaries” and thus have to be understood in social and professional contexts like other abstract concepts (Sadler, 2010, p. 9). In order to be able to use such criteria, students need to understand the concepts and practice until they can use the term appropriately, and explain and justify their judgements, which requires teachers’ guidance (Sadler, 2010).

In addition to teaching how to apply the rubric, teachers need to provide feedback on students’ self-assessment to enhance students’ abilities like the other skills for which they give feedback. The findings from the teachers’ interviews showed that the teacher's feedback is not directly related to their students’ self-assessment skills (their quality judgement). They provided feedback on the interpreting performance, their cause-analysis, their future improvement strategies, and their academic writing, but not their competence to self-assess. Students might guess whether their judgement on their performance was reasonable by comparing it to that of their teacher’s. This has a limitation in that it is possible only when the teachers’ feedback specifically includes the criteria so they can calibrate their judgement. Without a clear intention of fostering students’ skills in rubric application, however, like the case of explaining the rubric earlier, it appears unlikely that the teachers consciously provide feedback on the performance in a way that shows students how to apply the criteria.

The last condition is about giving students opportunities to practice self-assessment. However, the journal writing assignment took place only once in the semester in the classes included in this study. Interpreter teachers have many years of experience in teaching and/or

practicing interpreting in the field, and thus became expert examiners who can be intuitive in making a judgement (Tai, Ajjawi, Boud, Dawson, & Panadero, 2017). It is unlikely that students become capable of making a quality judgement on interpreting performance with a similar level of expertise, even upon completion of the program (Taras, 2008). Therefore, students need repeated self-assessment practice on various types of interpreting tasks because "making an evaluative judgement requires the activation of knowledge about quality in relation to a problem space" (Tai et al., 2017, p. 6).

Overall, the teachers' instruction in this study does not seem to be closely linked with self-assessment skill development. The findings about the student rubric application and the accuracy of their self-assessment appear to indicate that the students might not have developed the ability to self-assess their interpreting performance quality, and this suggests that the skill development status was related to the teachers' instruction to some extent.

5.4 Limitations

This study investigated three practical interpreting classes in one interpreter education program with a limited number of students. The inconsistency in the way self-assessment was incorporated and administered across the classes under the same unit sharing the same rubric, as well as across the program, led to the need to exclude some data and thus made the available data set even smaller. It is notoriously difficult in the discipline of interpreting studies to access large amounts of data. Also, since this is a case study within one institution, the findings about the students' use of the rubrics and performance on the assignment may be limited to this single case, and perhaps cannot be extended to other classes and other interpreting education programs in other institutions. In addition, it is hard to draw concrete conclusions that the students were not able to use certain criteria since they only wrote about one interpreting performance task, and I

did not have the opportunity to follow up with the students after analysing the data to further probe the reasons behind their decisions.

5.5 Suggestions for further research

There is still much to understand about interpreting students' abilities to self-assess. Therefore, further research is required. Firstly, we can start by investigating, using more direct means, such as observations, interviews or surveys, whether students experience difficulties or have confidence when making a quality judgement about their own interpreting performances. Secondly, we can identify issues that need to be addressed to support them in developing self-assessment skills, including using a set of criteria and the guidelines for applying them. Thirdly, we can identify which instructions are optimal for self-assessment, so that students can develop the skills to become competent. Lastly, the positive impact on autonomous and self-regulated learning in the interpreter education context of self-assessment skill development should be examined from a cognitive and psychological perspective, as it is the fundamental assumption about the benefit of self-assessment in learning.

CHAPTER 6: CONCLUSION

Using a mixed methods approach, this case study explored the competence of interpreting students in assessing their interpreting performance, the relationship between the competence in students' self-assessment and the results on their final assessments, and the relationship between students' ability in self-assessment and the instruction that they receive from their teachers. The findings suggest that students seem to have developed the ability to self-assess to some extent as their training progresses. For example, they are able to apply most of the criteria in the given rubric and identify specific constructs that affect the quality of an interpreting performance. However, it was observed that students tend to focus on only a few criteria, the concepts of which are relatively concrete, and avoid using abstract concepts. In addition, the students mainly discuss their mistakes, and this tendency appears to be reflected in their self-ratings which for the majority were an underestimation when compared to the scores attributed by their teachers. While the accuracy of their self-ratings showed a weak relationship with the outcomes on the final exam, this inverse relationship suggests that some students might have evaluated their performances against a higher standard and used the disappointment to motivate their learning.

The fundamental assumption about self-assessment is that it is beneficial for students in learning when they identify their strengths and weaknesses in a rather objective way, using criteria and evaluating their performances. In this regard, the findings seem to indicate that the students in this study have not developed some self-assessment skills.

The results seem to show a relationship with the way teachers provide instruction in self-assessment. The teachers in this study assumed that the students have a good understanding of the criteria and the ability to use them, which seems to have led them to focus on cause-analysis and strategic planning, which are regarded as subsequent phases of self-assessment in self-

regulated learning and reflective practice theories. Although the teachers understood the importance of self-assessment in interpreter training, they seemed not to be aware of the need to teach students how to self-assess.

This study has provided a picture of interpreting students' self-assessment abilities and the relationship between the ability and the outcomes on their final assessment as well as teachers' instruction in self-assessment, although it is limited to the specific context in which this study was conducted. In addition, the reasons for students' selective use of the criteria, as well as possible psychological factors that might affect self-assessment, need to be further explored, perhaps by listening to their own voices.

Nonetheless, the implications of this study are that some self-assessment skills may not be intuitively developed by students in the course of their interpreting education program and thus students may require more explicit guidance from their teachers to develop a high level of competence in self-assessment. Future research directions for self-assessment will need to explore a pedagogical approach to self-assessment skill development in interpreter education. As the fundamental purpose of self-assessment is to support interpreting students to become autonomous and self-regulated learners, identifying the link between self-assessment skill development and autonomous and self-regulated learning is essential.

References

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: Seven research-based principles for smart teaching* (1 ed.). San Francisco, CA: Jossey-Bass.

Andrade, H., & Du, Y. (2007). Student responses to criteriareferenced self-assessment. *Assessment & Evaluation in Higher Education, 32*(2), 159–181.

Andrade, H., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice, 27*(3), 3-13. doi:10.1111/j.1745-3992.2008.00118.x

Andrade, H., & Valtcheva, A. (2009). Promoting Learning and Achievement Through Self-Assessment. *Theory Into Practice, 48*(1), 12-19. doi:10.1080/00405840802577544

Austin, Z., & Gregory, P. A. M. (2007). Evaluating the Accuracy of Pharmacy Students' Self-Assessment Skills. *American Journal of Pharmaceutical Education, 71*(5), 1-8. doi:10.5688/aj710589

Ayres, L. (2008). Semi-Structured Interview. In L. M. Given (Ed.), *The SAGE Encyclopedia of Qualitative Research Methods* (pp. 810-811). Thousand Oaks: SAGE Publications, Inc.

Baker, M., & Maier, C. (2011). Ethics in intepreter & translator training. *The Interpreter and Translator Trainer, 5*(1), 1-14. doi:10.1080/13556509.2011.10798809

Barik, H. C. (1975/2002). Simultaneous Interpretation: Qualitative and Linguistic Data. In F. Pöchhacker & M. Shlesinger (Eds.), *The Interpreting Studies Reader* (pp. 79-91). New York: Routledge.

Bartłomiejczyk, M. (2006). Strategies of simultaneous interpreting and directionality. *Interpreting, 8*(2), 149-174. doi:10.1075/intp.8.2.03bar

Bartłomiejczyk, M. (2007). Interpreting Quality As Perceived by Trainee Interpreters. *The Interpreter and Translator Trainer, 1*(2), 247-267.

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university: what the student does* (4th ed.). Maidenhead: McGraw-Hill/Society for Research into Higher Education/Open University Press.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the Black Box: Assessment for Learning in the Classroom. *Phi Delta Kappan Magazine, 86*(1), 8-21. doi:10.1177/003172170408600105

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability(formerly: Journal of Personnel Evaluation in Education), 21*(1), 5-31. doi:10.1007/s11092-008-9068-5

Blenkinsopp, J., & Pajouh, M. S. (2010). Lost in translation? Culture, language and the role of the translator in international business. *Critical perspectives on international business, 6*(1), 38-52.

Bloom, B. S., Madaus, G. F., & Hastings, J. T. (1971). *Handbook on Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill.

Bontempo, K. M. (2012). *Interpreting by design : a study of aptitude, ability and achievement in Australian sign language interpreters.* (Ph.D), Australia : Macquarie University,

Boston, C. (2002). The Concept of Formative Assessment. *Practical Assessment, Research & Evaluation, 8*(9), Available online: http://PAREonline.net/getvn.asp?v=8&n=9.

Boud, D. (1995). *Enhancing learning through self assessment*. London: Kogan Page.

Boud, D., & Falchikov, N. (2007). Developing Assessment for Informing Judgement. In D. Boud & N. Falchikov (Eds.), *Rethinking Assessment for Higher Education: Learning for the Longer Term* (pp. 181-197). London: Routledge.

Boud, D., Keogh, R., & Walker, D. (1985). *Reflection: Turning Experience into Learning*. Oxon: Routledge.

Brown, A. (1987). Metacognition, executive control, self-regualtion, and other more mysterious mechanisms. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65-116). Hillsdale, N.J: L. Erlbaum Associates.

Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *The Sage handbook of research on classroom assessment* (pp. 367-393). Thousand Oaks, CA: Sage.

Bühler, H. (1986). Linguistic (Semantic) and Extra-linguistic (Pragmatic) Criteria for the Evaluation of Conference Interpretation and Interpreters. *Multilingua, 5*(4), 231-235.

Burns, R. B. (2000). *Introduction to Research Methods*: SAGE Publications.

Cassidy, S. (2007). Assessing ‘inexperienced’ students' ability to self‐assess: exploring links with learning style and academic personal control. *Assessment & Evaluation in Higher Education, 32*(3), 313-330. doi:10.1080/02602930600896704

Chappuis, S., & Chappuis, J. (2007). The best value in formative assessment. *Educational Leadership, 65*(4), 14-18.

Chiaro, D., & Nocella, G. (2004). Interpreters’ perception of linguistic and non-linguistic factors affecting quality: A survey through the World Wide Web. *Meta, 49*(2), 278-293.

Choi, J.-Y. (2004). A Metacognitive Approach to Evaluating Consecutive Interpretation for Novice Learners *Conference Interpretation and Translation, 6*(2), 169-184.

Choi, J.-Y. (2006). Metacognitive Evaluation Method in Consecutive Interpretation for Novice Learners. *Meta, 51*(2), 273-283.

Cohen, L., Manion, L., & Morrison, K. (2017). *Research Methods in Education*: Routledge.

Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative and Mixed Methods Approaches* (4 ed.). Thousand Oaks, CA: Sage.

Creswell, J. W., & Clark, V. L. P. (2011). *Designing and Conducting Mixed Methods Research*: SAGE Publications.

Dignath, C., Büttner, G., & Langfeldt, H. P. (2008). How can primary school students learn self-regulated learning strategies most effectively? A meta-analysis on self-regulation training programmes. *Educational Research Review, 3*, 101-129.

Doğan, A., Arumi Ribas, M., & Mora-Rubio, B. (2009). Metacognitive Tools in Interpreting Training: A Pilot Study. *Edebiyat Fakültesi Dergisi/Journal of Faculty of Letters, 26*(1), 69-84.

Dörnyei, Z. n. (2007). *Research methods in applied linguistics : quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.

Dunn, K. E., & Mulvenon, S. W. (2009). A Critical Review of Research on Formative Assessment: The Limited Scientific Evidence of the Impact of Formative Assessment in Education. *Practical Assessment, Research & Evaluation, 14*(7), 1-11.

Earl, L. M. (2005). *Thinking about purpose in classroom assessment : assessment for, as and of learning*. Deakin West, A.C.T.: ACSA.

Ericsson, K. A. (2004). Deliberate Practice and the Acquisition and Maintenance of Expert Performance in Medicine and Related Domains. *Academic Medicine, 79*(10), 70-81. doi:10.1097/00001888-200410001-00022

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The Role of Deliberate Practice in the Acquisition of Expert Performance. *Psychological Review, 100*(3), 363-406.

Evans, D. J. R., Zeun, P., & Stanier, R. A. (2014). Motivating student learning using a formative assessment journey. *Journal of Anatomy, 224*, 296-303. doi:10.1111/joa.12117

Falchikov, N., & Boud, D. (1989). Student Self-Assessment in Higher Education: A Meta-Analysis. *Review of Educational Research, 59*(4), 395-430.

Fernández, E. I. (2013). Unpacking delivery criteria in interpreting quality assessment. In D. Tsagari & R. van Deemter (Eds.), *Assessment issues in language translation and interpreting* (pp. 51-66). Frankfurt: Peter Lang.

Finlay, L. (2008). *Reflecting on 'Reflective practice'*. In. Retrieved from http://www.open.ac.uk/opencetl/sites/www.open.ac.uk.opencetl/files/files/ecms/web-content/Finlay-%282008%29-Reflecting-on-reflective-practice-PBPL-paper-52.pdf

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231-236). N.J: Erlbaum: Hillsdale.

Flavell, J. H. (1979). Metacognition and Cognitive Monitoring: A New Area of Cognitive-Developmental Inquiry. *American Psychologist, 34*(10), 906-911. doi:10.1037/0003-066X.34.10.906

Fowler, Y. (2007). Formative assessment: Using peer and self-assessment in interpreting training. In C. Wadensjö, B. E. Dimitrova, & A.-L. Nilsson (Eds.), *The Critical Link 4: Professionalisation of interpreting in the community. Selected papers from the 4th International Conference on Interpreting in Legal, Health and Social Service Settings* (pp. 253-262). Stockholm: John Bengamins Publishing Company.

Gerver, D. (1969/2002). The effects of Source Language Presentation Rate on the Performance of Simultaneous Conference Interpreters. In F. Pöchhacker & M. Shlesinger (Eds.), *The Interpreting Studies Reader* (pp. 55-66). New York: Routledge.

Gibbs, G. (1988). *Learning by doing: A guide to teaching and learning methods*. London: Further Education Unit.

Gibbs, G. (2010). *Using assessment to support student learning*: Leeds Metropolitan University.

Gibbs, G., & Simpson, C. (2004). Conditions Under Which Assessment Supports Students' Learning. *Learning and Teaching in Higher Education*(1), 3-31.

Gile, D. (2009). *Basic concepts and models for interpreter and translator training*. Amsterdam, Netherlands: John Benjamins Publishing Company.

Hale, S. B. (1997). The Interpreter on Trial: Pragmatics in Court Interpreting. In S. E. Carr, R. Roberts, A. Dufour, & D. Steyn (Eds.), *The Critical Link: Interpreters in the Community. Papers from the First International Conference on Interpreting in Legal, Health, and Social Service Settings (Geneva Park, Canada, June 1-4, 1995)* (pp. 201-211). Amsterdam: John Benjamins.

Hale, S. B. (2007). *Community Interpreting*. Basingstoke: Palgrave Macmillan Limited.

Han, C. (In press). A longitudinal quantitative investigation into the concurrent validity of self and peer assessment applied to English-Chinese bi-directional interpretation in an undergraduate interpreting course. *Studies in Educational Evaluation*. doi:https://doi.org/10.1016/j.stueduc.2018.01.001

Harris, B. (1990). Norms in Interpretation. *Target, 2*(1), 115-119.

Hartley, A., Mason, I., Peng, G., & Perez, I. (2003). *Peer- and Self-Assessment in Conference Interpreter Training*. Retrieved from http://www.llas.ac.uk/resourcedownloads/1454/hartley.rtf

Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, & K. Hakuta (Eds.), *Children development and education in Japan* (pp. 262-272). New York: Freeman.

Hild, A. (2014). The role of self-regulatory processes in the development of interpreting expertise. *Translation and Interpreting Studies, 9*(1), 128–149. doi:10.1075/tis.9.1.07hil

Hoffman, R. R. (1997). The cognitive psychology of expertise and the domain of interpreting. *Interpreting, 2*(1/2), 189-230.

Ivars, A. J., & Calatayud, D. P. (2013). Mindfulness training for interpreting students. *Lebende Sprachen, 58*(2), 341-365.

Jacobs, J. E., & Paris, S. G. (1987). Children's metacognition about reading: Issues in definition, measurement, and instruction. *Educational Psychologist, 22*(3-4), 255-278. doi:10.1080/00461520.1987.9653052

Johns, C. (1994). A philosophical basis for Nursing Practice. In C. Johns (Ed.), *The Burford NDU Model: Caring in Practice*. Oxford: Blackwell Scientific Publications.

Johns, C., & Freshwater, D. (2009). *Transforming nursing through reflective practice*: John Wiley & Sons.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher, 33*(7), 14-26. doi:10.3102/0013189X033007014

Kaefer, F., Roper, J., & Sinha, P. (2015). A Software-Assisted Qualitative Content Analysis of News Articles: Example and Reflections. *Forum: Qualitative Social Research, 16*(2). doi:10.17169/fqs-16.2.2123

Kim, Y.-H., & Chiu, C.-Y. (2011). Emotional Costs of Inaccurate Self-Assessments: Both Self-Effacement and Self-Enhancement Can Lead to Dejection. *Emotion, 11*(5), 1096-1104. doi:10.1037/a0025478

Kolb, D. (1984). *Experiential learning: Experience as a source and learning and development*. Englewood Cliffs, N.J.: Prentice-Hall.

Kostons, D., van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning & Instruction, 22*(2), 121-132. doi:10.1016/j.learninstruc.2011.08.004

Kurz, I. (2001). Conference Interpreting: Quality in the Ears of the User. *Meta, 46*(2), 394-409.

Laerd Statistics. (2015). Statistical tutorials and software guides. Retrieved from https://statistics.laerd.com/

Larrivee, B. (2000). Transforming Teaching Practice: Becoming the critically reflective teacher. *Reflective Practice, 1*(3), 293-307. doi:10.1080/713693162

Leach, L. (2012). Optional self-assessment: some tensions and dilemmas. *Assessment & Evaluation in Higher Education, 37*(2), 137-147. doi:10.1080/02602938.2010.515013

Lee, J. (2008). Rating Scales for Interpreting Performance Assessment. *The Interpreter and Translator Trainer, 2*(2), 165-184. doi:10.1080/1750399X.2008.10798772

Lee, S.-B. (2015). Developing an analytic scale for assessing undergraduate students' consecutive interpreting performances. *Interpreting, 17*(2), 226-254. doi:doi:10.1075/intp.17.2.04lee

Lee, Y. H. (2005). Self-assessment as an Autonomous Learning Tool in an Interpretation Classroom. *Meta, 50*(4).

Lee, Y. H. (2011). Comparing Self-assessment and Teacher's Assessment in Interpreter Training. *T&I Review, 1*, 87-111.

Lew, M. D. N., Alwis, W. A. M., & Schmidt, H. G. (2010). Accuracy of students' self-assessment and their beliefs about its utility. *Assessment & Evaluation in Higher Education, 35*(2), 135-156. doi:10.1080/02602930802687737

Li, X. (2018). Self-assessment as 'assessment as learning' in translator and interpreter education: validity and washback. *The Interpreter and Translator Trainer, 12*(1), 48-67. doi:10.1080/1750399X.2017.1418581

Liu, M., Chang, C.-c., & Wu, S.-c. (2008). Interpretation evaluation practices: Comparison of eleven schools in Taiwan, China, Britain, and the USA. *Compliation and Translation Review, 1*(1), 1-42.

Maguire, S., Evans, S. E., & Dyas, L. (2001). Approaches to Learning: A study of first-year geography undergraduates. *Journal of Geography in Higher Education, 25*(1), 95-107. doi:10.1080/03098260125539

McDonald, B., & Boud, D. (2003). The Impact of Self-assessment on Achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education: Principles, Policy & Practice, 10*(2), 209-220.

Moser-Mercer, B. (2008). Skill Acquisition in Interpreting. *The Interpreter and Translator Trainer, 2*(1), 1-28. doi:10.1080/1750399X.2008.10798764

Motta, M. (2016). A blended learning environment based on the principles of deliberate practice for the acquisition of interpreting skills. *The Interpreter and Translator Trainer, 10*(1), 133-149. doi:10.1080/1750399X.2016.1154347

Napier, J. (2004). Interpreting omissions: a new perspective. *Interpreting, 6*(2), 117-142.

Nicol, D. J., & Macfarlane‐Dick, D. (2006). Formative assessment and self‐regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199-218. doi:10.1080/03075070600572090

Nielsen, K. (2014). Self‐assessment methods in writing instruction: a conceptual framework, successful practices and essential strategies. *Journal of Research in Reading, 37*(1), 1-16. doi:10.1111/j.1467-9817.2012.01533.x

Olina, Z., & Sullivan, H. (2004). Student self-evaluation, teacher evaluation, and learner performance. *Educational Technology Research and Development, 52*(3), 5-22. doi:10.1007/BF02504672

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*(422). doi:10.3389/fpsyg.2017.00422

Panadero, E., & Alonso-Tapia, J. (2013). Self-Assessment: Theoretical and Practical Connotations. When It Happens, How Is It Acquired and What to Do to Develop It in Our Students. *Electronic Journal of Research in Educational Psychology, 11*(2), 551-576. doi:10.14204/ejrep.30.12200

Panadero, E., Brown, G. T. L., & Strijbos, J.-W. (2016). The Future of Student Self-Assessment: A Review of Known Unknowns and Potential Directions. *Educational Psychology Review, 28*(4), 803-830. doi:10.1007/s10648-015-9350-2

Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and selfefficacy: Four meta-analyses. *Educational Research Review, 22*, 74-98.

Pinazo, E. P. (2008). Self-Assessment in Teaching Interpreting. *TTR: Traduction, Terminologie, Rédaction: Etudes sur le Texte et Ses Transformations, 21*(1), 173-209. doi:10.7202/029690ar

Pöchhacker, F. (2002). Researching interpreting quality. In G. Garzone & M. Viezzi (Eds.), *Interpreting in the 21st century : challenges and opportunities* (pp. 95-106). Amsterdam: J. Benjamins.

Pöchhacker, F. (2012). INTERPRETING QUALITY: GLOBAL PROFESSIONAL STANDARDS? In R. W (Ed.), *Interpreting in the Age of Globalization: Proceedings of the 8th National Conference and International Forum on Interpreting* (pp. 305-318). Beijing: Foreign Language Teaching and Research Press.

Pöchhacker, F. (2015). *Routledge encyclopedia of interpreting studies*. New York: Taylor & Francis.

Pöchhacker, F. (2016). *Introducing interpreting studies* (Second edition. ed.). New York: Routledge.

Quinn, F. M. (2000). Reflection and reflective practice. In C. Davies, L. Finlay, & A. Bullman (Eds.), *Changing practice in health and social care*: SAGE Publications.

Riccardi, A. (2002). Evaluation in Interpretation: Macrocriteria and Microcriteria. In E. Hung (Ed.), *Teaching Translation and Interpreting 4: Building Bridges* (pp. 115-126). Amsterdam: John Benjamins.

Rivers, W. P. (2001). Autonomy at All Costs: An Ethnography of Metacognitive Self-Assessment and Self-Management among Experienced Language Learners. *Modern Language Journal, 85*(2), 279-290. doi:10.1111/0026-7902.00109

Ross, J. A. (2006). The Reliability, Validity, and Utility of Self-Assessment. *Practical Assessment, 11*(10), 1-13.

Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002). Student Self-Evaluation in Grade 5-6 Mathematics Effects on Problem- Solving Achievement. *Educational Assessment, Evaluation and Accountability(formerly: Journal of Personnel Evaluation in Education), 8*(1), 43-58. doi:10.1207/S15326977EA0801_03

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119-144. doi:10.1007/BF00117714

Sadler, D. R. (2010). Beyond Feedback: Developing Student Capability in Complex Appraisal. *Assessment & Evaluation in Higher Education, 35*(5), 535-550. doi:10.1080/02602930903541015

Sawyer, D. (2004). *Fundamental Aspects of Interpreter Education : Curriculum and Assessment*. Amsterdam: John Benjamins.

Schön, D. (1983). *The Reflective Practitioner: How Professionals Think in Action*. London: Temple Smith.

Schön, D. (1987). *Educating the reflective practitioner*. San Francisco Jossey-Bass.

Schunk, D. H. (2003). Self-efficacy for reading and writing: Influence of modeling, goal setting, and self-evaluation. *Reading and Writing Quarterly, 19*, 159-172.

Scriven, M. (1967). The Methodology of Evaluation. *AERA Monograph Series on Evaluation, 1*, 39-83.

Siegesmund, A. (2017). Using self-assessment to develop metacognition and self-regulated learners. *FEMS Microbiology Letters, 364*(11), 1-4. doi:doi: 10.1093/femsle/fnx096

Silver, N. (2011). Reflective pedagogies and the metacognitive turn in college teaching. In M. Kaplan, N. Silver, D. LaVaque-Manty, & D. Meizlish (Eds.), *Using reflection and*

metacognition to improve student learning: Across the disciplines, across the academy (pp. 1-17). Herndon, UNITED STATES: Stylus Publishing.

Slatyer, H. (2015). *Multilingual interpreter education curriculum design and evaluation.* (PhD), Macquarie University, Sydney, Australia.

Slatyer, H., & Chesher, T. (2007). Talking about accuracy. Interpreters' understanding of a key principle of professional ethics. In F. Pöchhacker, A. L. Jakobsen, & I. M. Mees (Eds.), *Interpreting Studies and Beyond* (pp. 139-152). Frederiksberg C: Samfundslitteratur.

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2017). Developing evaluative judgement: enabling students to make decisions about the quality of work. *High Educ*. doi:https://doi.org/10.1007/s10734-017-0220-3

Taras, M. (2005). Assessment – summative and formative – some theoretical reflections. *British Journal of Educational Studies, 53*(4), 466-478. doi:10.1111/j.1467-8527.2005.00307.x

Taras, M. (2008). Issues of power and equity in two models of self-assessment. *Teaching in Higher Education, 13*(1), 81-92. doi:10.1080/13562510701794076

Taras, M. (2010). Student self-assessment: processes and consequences. *Teaching in Higher Education, 15*(2), 199-209. doi:10.1080/13562511003620027

Tashakkori, A., & Teddlie, C. (2010). Putting the Human Back in "Human Research Methodology": The Researcher in Mixed Methods Research. *Journal of Mixed Methods Research, 4*(4), 271-277. doi:10.1177/1558689810382532

Tsui, A. B. M. (2009). Distinctive qualities of expert teachers. *Teachers and Teaching, 15*(4), 421-439. doi:10.1080/13540600903057179

Wiliam, D. (2000, August). *Integrating formative and summative functions of assessment.* Paper presented at the Working group 10 of the International Congress on Mathematics Education, Makuhari, Tokyo.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*(1), 3-14. doi:https://doi.org/10.1016/j.stueduc.2011.03.001

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education, 11*, 46-65.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). NJ: Erlbaum: Mahwah.

Wininger, R. S. (2005). Using your tests to teach: Formative summative assessment. *Teaching Psychology, 32*(2), 164-166.

Winne, P. H. (2011). A Cognitive and Metacognitive Analysis of Self-Regulated Learning. In D. H. Schunk & B. Zimmerman (Eds.), *Handbook of self-regulation of learning and performance* (pp. 15-32). New York: Taylor and Francis.

Winne, P. H. (2017). Cognition and metacognition within self-regulated learning. In P. A. Alexander, D. H. Schunk, & J. A. Greene (Eds.), *Handbook of Self-Regulation of Learning and Performance*: Routledge Handbooks Online.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. . In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). NJ:Erlbaum: Mahwah.

Wu, F. (2013). How do we assess students in the interpreting examinations? In D. Tsagari & R. van Deemter (Eds.), *Assessment issues in language translation and interpreting* (pp. 15-34). Frankfurt: Peter Lang.

Wu, Z. (2016). Towards understanding interpreter trainees' (de)motivation: An exploratory study. *Translation & Interpreting, 8*(2), 13-25. doi:10.12807/ti.108202.2016.a02

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64-70.

Zimmerman, B. J., & Kitsantas, A. (1999). Acquiring writing revision skill: Shifting from process to outcome self-regulatory goals. *Journal of Educational Psychology, 99*, 241–250.

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regualtion. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of Metacognition in Education* (pp. 299-315). New York, NY: Routledge.

Zimmerman, B. J., & Schunk, D. H. (2011). Self-Regulated Learning and Performance. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of Self-Regulation of Learning and Performance*. New York: Routledge.

Zwischenberger, C. (2010). Quality criteria in simultaneous interpreting: An international vs. a national view. *The Interpreters' Newsletter, 15*, 127–142.

Appendix 1 Ethics Committee Approval

Date: 30 October 2017

Dear Dr Slatyer,

Re: "Interpreting students' ability in self-assessment practices" (5201700979)

Thank you very much for your response. Your response has addressed the issues raised by the Faculty of Human Sciences Human Research Ethics Sub-Committee and approval has been granted, effective 30th October 2017. This email constitutes ethical approval only.

This research meets the requirements of the National Statement on Ethical Conduct in Human Research (2007). The National Statement is available at the following web site:

https://www.nhmrc.gov.au/book/national-statement-ethical-conduct-human-research

The following personnel are authorised to conduct this research:

Dr Helen Slatyer

Miss Eunjin Heo

NB. STUDENTS: IT IS YOUR RESPONSIBILITY TO KEEP A COPY OF THIS APPROVAL EMAIL TO SUBMIT WITH YOUR THESIS.

Please note the following standard requirements of approval:

1. The approval of this project is conditional upon your continuing compliance with the National Statement on Ethical Conduct in Human Research (2007).

2. Approval will be for a period of five (5) years subject to the provision of annual reports.

Progress Report 1 Due: 30th October 2018

Progress Report 2 Due: 30th October 2019

Progress Report 3 Due: 30th October 2020

Progress Report 4 Due: 30th October 2021

Final Report Due: 30th October 2022

NB. If you complete the work earlier than you had planned you must submit a Final Report as soon as the work is completed. If the project has been discontinued or not commenced for any reason, you are also required to submit a Final Report for the project.

Progress reports and Final Reports are available at the following website:

http://www.research.mq.edu.au/current_research_staff/human_research_ethics/resources

3. If the project has run for more than five (5) years you cannot renew approval for the

project. You will need to complete and submit a Final Report and submit a new application for the project. (The five year limit on renewal of approvals allows the Sub-Committee to fully re-review research in an environment where legislation, guidelines and requirements are continually changing, for example, new child protection and privacy laws).

4. All amendments to the project must be reviewed and approved by the Sub-Committee before implementation. Please complete and submit a Request for Amendment Form available at the following website:

http://www.research.mq.edu.au/current_research_staff/human_research_ethics/managing_approved_research_projects

5. Please notify the Sub-Committee immediately in the event of any adverse effects on participants or of any unforeseen events that affect the continued ethical acceptability of the project.

6. At all times you are responsible for the ethical conduct of your research in accordance with the guidelines established by the University. This information is available at the following websites:

http://www.mq.edu.au/policy

http://www.research.mq.edu.au/current_research_staff/human_research_ethics/managing_approved_research_projects

If you will be applying for or have applied for internal or external funding for the above project it is your responsibility to provide the Macquarie University's Research Grants Management Assistant with a copy of this email as soon as possible. Internal and External funding agencies will not be informed that you have approval for your project and funds will not be released until the Research Grants Management Assistant has received a copy of this email.

If you need to provide a hard copy letter of approval to an external organisation as evidence that you have approval, please do not hesitate to contact the Ethics Secretariat at the address below.

Please retain a copy of this email as this is your official notification of ethics approval.

Yours sincerely,

Dr Naomi Sweller

Chair

Faculty of Human Sciences

Human Research Ethics Sub-Committee

Appendix 2 Participant Consent form

Student participants

Department of Linguistics
Faculty of Human Sciences
MACQUARIE UNIVERSITY NSW 2109
Phone: +61 (2) 9850-8651
Fax: +61 (2) 9850 9199
Email: helen.slatyer@mq.edu.au

Chief Investigator's / Supervisor's Name & Title: Dr Helen Slatyer

Participant Information and Consent Form (Students)

Name of Project: Interpreting students' ability in self-assessment

You are invited to participate in a study of Interpreting students' ability in self-assessment. The purpose of the study is to explore interpreting students' ability to assess their own interpreting performance.

The study is being conducted by Eunjin Heo (eunjin.heo@hdr.mq.edu.au) to meet the requirements of the Master of Research under the supervision of Dr Helen Slatyer (+61 (2) 9850-8651, helen.slatyer@mq.edu.au) of the Department of Linguistics.

If you decide to participate, you will be asked to allow the researchers to access to the reflective journal you submitted as an assignment, the grade and feedback you received on the reflective journal and your performance from your teacher, and your final exam grade. You will also be asked to mark your own performance about which you wrote the reflective journal. The time it takes you to mark your performance will be variable, but shouldn't take you more than 45 minutes. There are no known risks associated with participation in this study.

Any information or personal details gathered in the course of the study are confidential, except as required by law. No individual will be identified in any publication of the results. Only the student researcher and her supervisor will have access to the data. All data will be assigned a code to ensure anonymity and pseudonyms will be used in reports of the results. A summary of the results of the data can be made available to you on request via email to Eunjin Heo.

Participation in this study is entirely voluntary: you are not obliged to participate and if you decide to participate, you are free to withdraw at any time without having to give a reason and without consequence. Your participation has no bearing on your study in the Graduate Diploma of Translation or Interpreting or Masters of Translation and Interpreting Studies.

I, *(participant's name)* have read *(or, where appropriate, have had read to me)* and understand the information above and any questions I have asked have been answered to my

satisfaction. I agree to participate in this research, knowing that I can withdraw from further participation in the research at any time without consequence. I have been given a copy of this form to keep.

Participant's Name:_____
(Block letters)

Participant's Signature: _____ Date:_____

Investigator's Name:_____
(Block letters)

Investigator's Signature:_____ __ Date:_____

The ethical aspects of this study have been approved by the Macquarie University Human Research Ethics Committee. If you have any complaints or reservations about any ethical aspect of your participation in this research, you may contact the Committee through the Director, Research Ethics & Integrity (telephone (02) 9850 7854; email ethics@mq.edu.au). Any complaint you make will be treated in confidence and investigated, and you will be informed of the outcome.

(PARTICIPANT'S COPY)

Teaching staff participants

Department of Linguistics
Faculty of Human Sciences
MACQUARIE UNIVERSITY NSW 2109
Phone: +61 (2) 9850-8651
Fax: +61 (2) 9850 9199
Email: helen.slatyer@mq.edu.au

Chief Investigator's / Supervisor's Name & Title: Dr Helen Slatyer

Participant Information and Consent Form (Teachers)

Name of Project: Interpreting students' ability in self-assessment

You are invited to participate in a study of Interpreting students' ability in self-assessment. The purpose of the study is to explore interpreting students' ability to assess their own interpreting performance themselves.

The study is being conducted by Eunjin Heo (eunjin.heo@hdr.mq.edu.au) to meet the requirements of the Master of Research under the supervision of Dr Helen Slatyer (+61 (2) 9850-8651, helen.slatyer@mq.edu.au) of the Department of Linguistics.

If you decide to participate, you will be asked to provide information about the assessments of students enrolled TRAN882 or TRAN822 who have consented to participate in the study: students' reflective journals, the grade (where available) and feedback that you have given for the reflective journals and the breakdown of students' grades on the final exam grade. You will also be invited to participate in a one-on-one interview about self-assessment practices in these two units with Eunjin Heo, the student researcher. The interview will be audio-recorded. There are no known risks associated with participation in this study.

Any information or personal details gathered in the course of the study are confidential, except as required by law. No individual will be identified in any publication of the results. Only the student researcher and her supervisor will have access to the data. All data will be assigned a code to ensure anonymity and pseudonyms will be used in reports of the results. A summary of the results of the data can be made available to you on request via email to Eunjin Heo.

Participation in this study is entirely voluntary: you are not obliged to participate and if you decide to participate, you are free to withdraw at any time without having to give a reason and without consequence.

I, *(participant's name)* have read *(or, where appropriate, have had read to me)* and understand the information above and any questions I have asked have been answered to my satisfaction. I agree to participate in this research, knowing that I can withdraw from further participation in the research at any time without consequence. I have been given a copy of this form to keep.

Participant's Name:_____
(Block letters)

Participant's Signature: _____ Date:_____

Investigator's Name:_____
(Block letters)

Investigator's Signature:_____ __ Date:_____

The ethical aspects of this study have been approved by the Macquarie University Human Research Ethics Committee. If you have any complaints or reservations about any ethical aspect of your participation in this research, you may contact the Committee through the Director, Research Ethics & Integrity (telephone (02) 9850 7854; email ethics@mq.edu.au). Any complaint you make will be treated in confidence and investigated, and you will be informed of the outcome.

(PARTICIPANT'S COPY)

Appendix 3 The format of the interpreting tasks used for self-assessment in the units in the research

Unit	Mode	Direction	Length of passage
IP3	Consecutive	Korean into English	Two sets of 3- minute passage
IP2	Consecutive	LOTE into English, English into LOTE	Two sets of 2-minuite passage for each language direction

Appendix 4 Semi-structured interview questions

Semi-structured interview questions

Self-reflective Journal Writing Assignment

1. What interpreting task do students perform to write a reflective journal?
    a. What is the interpreting mode (consecutive, dialogue)?
    b. How long is the task?
2. What is the requirement for the assignment?
    a. Guideline for the assignment (e.g. word count, structure, due date)
    b. Are they required to give their interpreting performance a score?
    c. What is the weight of the grade for the assignment?
    d. How many times are students required to write a reflective journal on their performance for the unit?
3. How do you give students feedback on their reflective journal?
    a. What does your feedback include?
        i. Performance analysis
        ii. Interpreting performance quality
        iii. Journal writing skills
    b. Written
    c. Length
4. What is the purpose of giving an assignment to write a reflective journal?

Self-assessment Instruction

1. Could you please explain how you introduce self-assessment to your students?
    a. When, how often, and how much time is allocated for the instruction?
    b. Instruction material (e.g. description, rubric, samples)
2. Besides the reflective journal writing assignment, is there a way to give feedback on students' self-assessment?

Rubric for performance assessment

1. Do you provide a rubric to students to use for their self-assessment?
2. When do you provide your rubric to students?
3. How do you explain how to use the rubric?
    a. Do you explain each item in the rubric?
    b. How long does it take to explain the rubric?
4. When giving in-class feedback for students' performance, do you use terminology in the rubric?
5. Do you use the same rubric that you provide to students for exams?
    a. Do you tell students what rubric you use for exams?