

Correlating Self-Efficacy with Self-Assessment in an Undergraduate Interpreting Classroom: How Accurate can Students be?

JING LIU

Anhui Normal University

Received: 11 March 2020 / Accepted: 15 June 2021

DOI: 10.30827/portalin.v0i36.13897

ISSN paper edition: 1697-7467, ISSN digital edition: 2695-8244

ABSTRACT: The current paper intends to explore whether there are significant correlations between students' self-efficacy and their self-assessment accuracy and how the former mediates the latter. Framed within an undergraduate interpreting classroom in China, which shares similar pedagogical aims with general foreign language courses, a total of 53 senior students completed an Interpreting Self-Efficacy (ISE) Scale before self-assessing their English-Chinese consecutive interpreting performance. Spearman correlation tests were employed to investigate the correlations between students' ISE level and their self-assessment accuracy, compared with the teacher's marks. Although ISE and self-assessment accuracy were positively correlated, the relation was not significant. Medium to low level ISE could only vaguely predict students' self-assessment performance, but students were capable of accurate self-assessment regardless of their ISE level. This justifies more rigorous reflection on self-regulated learning enabled by accurate self-assessment in language classrooms, which is simultaneously informed by multiple social and psychological variables experienced by individual learners, such as self-efficacy.

Key words: interpreting self-efficacy, self-assessment, self-assessment accuracy, consecutive interpreting, Chinese students.

Correlación de la autoeficacia con la autoevaluación en un aula de interpretación de pregrado: Que tan exactos pueden ser los estudiantes?

RESUMEN: El artículo actual tiene la intención de explorar si existen correlaciones significativas entre la autoeficacia de los estudiantes y la precisión de su autoevaluación y cómo la primera interviene en la segunda. Enmarcado dentro de un aula de interpretación de pregrado en China, que comparte objetivos pedagógicos similares con cursos generales de idiomas extranjeros, un total de 53 estudiantes de último año completaron una escala de autoeficacia en la interpretación (ISE) antes de autoevaluar su desempeño de interpretación consecutiva inglés-chino. Se emplearon pruebas de correlación de Spearman para investigar las correlaciones entre el nivel de ISE de los estudiantes y la precisión de su autoevaluación, en comparación con las calificaciones del maestro. Aunque ISE y la precisión de la autoevaluación se correlacionaron positivamente, la relación no fue significativa. El ISE de nivel medio a bajo solo podía predecir vagamente el desempeño de la autoevaluación de los estudiantes, pero los estudiantes eran capaces de realizar una autoevaluación precisa independientemente de su nivel de ISE. Esto justifica una reflexión más rigurosa sobre el aprendizaje autorregulado gracias a una autoevaluación precisa en las aulas de idiomas, que está informada simul-

táneamente por múltiples variables sociales y psicológicas que experimentan los alumnos individuales, como la autoeficacia.

Palabras clave: autoeficacia de interpretación, autoevaluación, precisión de autoevaluación, interpretación consecutiva, estudiantes chinos.

1. INTRODUCTION

One of the ultimate aims of education is to enable students to better judge their own work (Boud et al., 2013), and self-assessment (SA) has been widely perceived as a powerful tool in this regard, as witnessed by its innovative implementation in various classrooms (Boud, 1995; Falchikov, 2004). By far, SA research has been dominated by discussion of its significance, validity and best practice (Huang, 2015; Hung, 2019), overshadowing the role played by learner variables, which is a prominent research area in second language acquisition (de Bot et al., 2005; Dörnyei, 2006). Among a long list of examined variables, self-efficacy stands out in a plethora of research (Bandura, 1994; Ivars et al., 2014; Wang et al., 2014). To ensure maximum investment from students in any course, it is of paramount importance to boost and maintain their self-efficacy, as this psychological trait directly influences how students respond to challenges (Bandura & Schunk, 1981). Moreover, self-efficacy is reported to inform many aspects, including learning styles and strategies, students' performance, etc., with SA emerging as one frequent topic (Cassidy, 2007; Chemers et al., 2001). Research abounds regarding the driving power of SA to self-efficacy (e.g., Andrade et al., 2009; Brown & Harris, 2013; Kissling, E. & O' Donnell, 2015; McMillan & Hearn, 2008; Panadero et al., 2012, 2013, 2017; Reybroeck et al., 2017), but studies on the reverse direction are scarce. Limited research reports divergent and inconclusive results as to whether and how self-efficacy informs SA (Chemers et al., 2001; Ivars et al., 2014; Shaban et al., 2016) despite the increasing recognition that accurate SA leads to accurate task selection, thus conducive to self-regulated learning and improved learning outcomes (Brown & Harris, 2013; McMillan & Hearn, 2008; Raaijmakers et al., 2019).

It is contended that self-efficacy should be context-specific (Pajares, 1996, p. 547) whereas SA is assumed to be context-sensitive in different disciplines. While the research in self-efficacy proliferates across multiple disciplines (Chan & Lam, 2010; Shaban et al., 2016; Wang et al., 2014), similar endeavours, especially empirical efforts correlating self-efficacy with SA, are scant in the context of interpreting classrooms (Li, 2018). It should be noted that undergraduate interpreting courses in Chinese context are more similar to regular foreign language courses, with the aim of reinforcing bilingual proficiency of students rather than cultivating professional interpreters and translators (Cai & Dong, 2015; Yan et al., 2018). For these students, inaccurate SA might result in inappropriate learning tasks, in turn undermining the monitoring and management of their own language learning (Andrade & Valtcheva, 2009; Kostons et al., 2012). As a result, studies into the factors moderating SA and its accuracy are all the more pertinent.

Although it is acknowledged that learner variables, including motivation, anxiety etc., mediate SA in one way or another (MacIntyre et al., 1997; Shaban et al., 2016), there has not been much research centred on accuracy explicitly investigating the exact link. The current study, thus, intends to delve into this area by exploring whether and how self-efficacy,

in particular, self-efficacy in interpreting, mediates SA accuracy. It is postulated that the higher self-efficacy students have, the more accurate their SA will be. To achieve this, in an undergraduate interpreting classroom, students were asked to assess their own consecutive interpretation after filling out a questionnaire measuring their interpreting self-efficacy. Certain patterns were recognised, and conclusions reached after the data were quantitatively analysed and qualitatively described. With this, the study hopes to inspire more discussion on mediating factors of SA accuracy in language learning from the perspective of learner variables and to help untangle the complexity in the classroom environment in which both teachers and students work (Panadero et al., 2016).

2. SELF-ASSESSMENT

2.1. Self-assessment in interpreting education

SA is the ability of a student to judge his or her own performance so as to make decisions about oneself and one's abilities based on criteria co-developed by students and teachers (Boud, 1995). In higher education, it is recognised as an influential instrument to enhance students' self-regulation and power in the learning process (Bourke, 2017; Chen, 2008).

Evidence-based studies of SA in translation and interpreting classrooms are still limited (Li, 2018). Interpreting courses aim at preparing students for professional interpreter careers (Lee, 2015; Postigo Pinazo, 2008), thus including a huge amount of exercises, external evaluation and self-assessment of student performance. The modules in a typical undergraduate interpreting course often comprise training on note-taking skills, memory skills, bi-directional consecutive interpreting, sometimes coupled with sight translation (Li, 2018; Postigo Pinazo, 2008).

Notwithstanding the highly demanding course design, in Chinese context, interpreting education is germane to foreign language education in multiple aspects (Yan & Wang, 2012; Yan et al., 2018). Specifically, interpreting courses in many undergraduate programs are part of the overall foreign language programs (Pan & Yan, 2012). Interpreting training is compulsory for almost all foreign language majors, the majority of whom are English majors. These students receive interpreting courses for one or two semesters, the main purpose of which is to enhance their second language competence like any other standardised language courses (Cai & Dong, 2015).

Meanwhile, the practice-oriented nature of interpreting courses, be they in China or elsewhere, determines a disproportionate overemphasis on appropriate criteria and accuracy when it comes to SA (e.g., Han & Riazi, 2018; Lee, 2015; Stauffer, 2011). Because without comprehensive criteria for assessing performance or an accurate judgement of one's own capability, it is nearly impossible for students to improve their weak points. It is noteworthy, however, that across consecutive, sight and sign language interpreting courses, the SA research generally targets advanced learners (e.g., Han & Riazi, 2018; Li, 2018; Stauffer, 2011), with the assumption that they are already capable of assessing themselves before taking the particular courses. Such design naturally invites suspicion because the representation of the data and the heterogeneous characters of the learners are left out.

2.2. Self-assessment accuracy

To support valid uses, an assessment is expected to be accurate and reliable in portraying a learner's achievement or performance (Brown et al., 2015). Therefore, one important focus of SA research regards the accuracy. Common practice is to compare students' marks with those of teachers, peers, external raters or standardised tests (e.g., Brown et al., 2015; Stauffer, 2011), holding teachers' judgement as the benchmark most of the time. Although existing research has identified significant correlation between SA and teacher assessment (Brown et al., 2015; Cassidy, 2007; Chen, 2008; Li, 2018), much of it elaborates on the tendency of students' under-reporting, especially for students from Chinese backgrounds (Chen, 2008; Li, 2018). To improve the consistency between SA and teacher assessment, various suggestions for future research have been made. For example, it is strongly recommended that before formal SA, training through a workshop or a tutorial should be organised as a necessary step in developing students' understanding of the criteria and enhancing the quality of SA (Chen, 2008). Researchers are advised to set clear criteria, train students, avoid reliability pitfalls and so on if they intend to compare SA marks with teacher marks (Brown et al., 2015; Han & Riazi, 2018).

Even so, the legitimacy of holding teacher marks as the golden standard has been called into question (Brown et al., 2015), because the skills of teachers and their ability to provide quality assessment vary widely. Nevertheless, given the fact that in most cases the same teachers also assess students in the formal examination, consistency can partially be guaranteed. Besides, unless SA scores are part of a summative test, the priority of SA for students, as Brown, Andrade and Chen (2015) argue, is more on their internal reflection concerning the whole learning process than on how well they do in assessing their own performance. In the same vein, it is accentuated that to serve as an effective assessment tool, SA should not simply depend on students' ability to mark themselves according to certain criteria, nor on the similarity between student and teacher marks. Rather, effectiveness should be judged according to the extent of the students' development during the process. That is, in addition to statistical analysis of self and teacher marks, how students apply the criteria to assessment, which indicates their understanding of the criteria, should not be overlooked (Orsmond et al., 1997).

However, if students are to appropriately monitor their learning process and adjust learning strategies to achieve autonomous learning, a certain level of SA accuracy is still required. But by far the findings in related research have not been consistent. Some research observes that students with poorer competence tend to over-report scores while more competent students are more stringent raters (Dragemark, 2006, p. 178); whereas better translation performance from trainee students accompanies better SA as these students are more aware of their shortcomings and how to improve their translation (Fernández & Zabalbeascoa, 2012). In contrast, in interpreting classrooms, advanced students are less likely to be accurate than beginning students (Stauffer, 2011). It is evident that more research is needed to determine patterns in SA accuracy as well as potential factors at play.

2.3. Variables mediating self-assessment accuracy

A list of variables has been proposed for closer inspection as to whether they moderate SA accuracy, including training, cultural background etc. (Brown et al., 2015). But much research stops short of touching upon these issues rather than delving into their depth. One exception is Hung (2019), who narrows down to the impact of training on SA accuracy. In a college English class, students practice SA of their oral English five times after receiving the teacher's instructions, discussions over assessment criteria, pilot SA and feedback from the teacher. It is observed that students' SA accuracy improves tremendously after repeated practice. With clear guidance and feasible design (i.e., students use cell phones to record their oral response), the methods in this study can be replicated in other classrooms with ease. However, it is argued that even inexperienced students are capable of accurate SA if they are equipped with enough self-efficacy (Cassidy, 2007). It is hypothesised that students who expect to obtain good grades and anticipate they will complete the task effectively also show positive self-efficacy beliefs. Reflected in SA marks, it can be expected that more efficacious students will mark themselves higher than less efficacious peers, leading to more SA accuracy if their self-perception is in line with teachers' observations. But this correlation is merely a suggestion yet to be confirmed.

When self-efficacy is pinpointed, its moderating role in SA is mixed. It is reported by some research that more efficacious students tend to over-report (Shaban et al., 2016) while others observe no significant correlations (Ng & Earl, 2008). To further complicate the picture, it is discovered that first-year college students become more accurate self-assessors with more self-efficacy (Chemers et al., 2001). But several pitfalls need to be considered. The students under scrutiny came from various academic backgrounds; the SA questions and teacher assessment were vague and obscure. Given a lack of focus and context, more validation is required to corroborate the findings. Similarly, Cassidy (2007) warns the multifaceted nature of SA requires a more cautious approach covering potential aspects that might have some influence. It should be noted that the contexts of aforementioned research were not in language education. Therefore, the findings should be interpreted with caution. Plagued by a dearth of research on the effects of self-efficacy on SA accuracy in language education, more vigorous investigations are required to clarify the intricate dynamics.

Besides self-efficacy, other variables are also believed to moderate SA accuracy. For example, anxiety, an affective factor in direct contrast to self-efficacy, is reported to systematically bias students' perceptions of competence, although the anecdotal belief holds that it does not inform SA accuracy (MacIntyre et al., 1997). That is, more anxious students tend to under-report while less anxious students over-report. Investigations into Chinese students self-assessing oral English tasks support this finding (Chen, 2008), as modesty, seen as a virtue in Chinese society, might be one of the reasons behind under-reporting. This is a meaningful step forward concerning the function of psychological traits combined with social-cultural variables in informing SA.

A similar trend is observed in the limited empirical studies of SA in interpreting. Han and Riazi (2018) take a quantitative approach in a longitudinal study to examine how SA accuracy evolves through practice. Although students' SA accuracy improves over time after repeated training, in English to Chinese consecutive interpreting they are more accurate than in Chinese to English interpreting, revealing a unique pattern regarding the role of language

directionality in affecting SA. This can be partly explained by comparing students' mother tongue (Chinese) and their foreign/target language (English), which affects their comprehension and assessment. Meanwhile, self-efficacy is observed to mediate performance among interpreting students, which will be reflected in their SA marks, but not without certain prerequisites. Self-efficacy is reported to predict students' performance when they are equipped with good linguistic proficiency (Ivars et al., 2014) or with high interpreting self-efficacy (Lee, 2018). It seems that self-efficacy can only improve performance among competent students while less competent ones cannot benefit at all. But as the next section will show, there is neither definite nor direct causal relations between self-efficacy and interpreting performance. Besides, in the learning process, a number of variables are active and more research is needed to clarify their relationship.

3. SELF-EFFICACY

Self-efficacy is defined as people's judgement of their capabilities to complete a specific task with the skills they possess (Bandura, 1997). It is a belief that influences virtually every sphere of people's lives (Ivars et al., 2014, p. 168) and determines how people feel, think, motivate themselves and behave (Bandura, 1994). Individuals with high self-efficacy invest more effort and persist longer than those with low self-efficacy, especially when they face setbacks (Chan & Lam, 2010). Defined as a task-specific attribute (Wang et al., 2014), self-efficacy has been examined in various contexts, among which language learning, particularly foreign language learning, stands out as one of the most heavily researched areas (e.g., Cassidy, 2007; Chan et al., 2010; Liu, 2018; Wang et al., 2014). Relevant research elaborates on the measurement of self-efficacy among English language learners in China (Wang et al., 2014) and the impact of different feedback modes on Chinese students' self-efficacy in English vocabulary acquisition (Chan et al., 2010). Self-efficacy has been closely correlated with many variables and the relationships are never straightforward. When self-efficacy is examined in translation and interpreting classrooms, researchers are generally interested in how it impacts students' performance and how it can be measured accurately (Ivars et al., 2014; Bolaños-Medina et al., 2018). Researchers either borrow self-efficacy scales from other disciplines or construct their own when measuring self-efficacy, with no consensus reached (e.g., Cassidy, 2007; Bontempo & Napier, 2011; Wang et al., 2014).

Nonetheless, Lee (2014) is recognised as the first to have constructed and validated a self-efficacy scale specifically targeting undergraduates in consecutive interpreting. The scale, a self-report questionnaire, is named the Interpreting Self-Efficacy (ISE) scale, which defines ISE as one's self-beliefs in interpreting competence. By measuring students' ISE, teachers can predict the performance of students facing an interpreting task. However, one drawback of the scale comes from the generalisation of results interpretation. Although it is claimed that the higher scores students receive, the more efficacious they are, researchers and students still have no clear referential benchmark. Also, the scale appears to have loaded statements that might interfere with obtaining accurate assessment. Since the social response bias is common in self-report questionnaires (Paulhus, 1991), the authenticity of the data gathered might be called into question.

In a follow-up study (Lee, 2018), the ISE scale is employed to predict the actual performance of students by triangulating different sources of assessment from the teacher, peer and external raters. It is found that ISE is more closely correlated with the performance among students with relatively high ISE levels, indicating that teachers can assess students' expected performance and task difficulty with more confidence, but only when they bear in mind there is no direct impact from one on the other. Lee particularly points out the potential interplay between ISE and other psychological or social-cultural factors in the prediction of student performance. It is quite interesting to note that among the factors at play, student SA is frequently cited (e.g., Andrade et al., 2009; Panadero et al., 2017). Multiple studies explore how SA influences student self-efficacy, but seldom has research looked into the reversed direction, namely, how self-efficacy might moderate the former.

To recapitulate, the common practice of measuring SA accuracy is by comparing student marks with teacher marks, despite some suspicion of its validity. Second language acquisition and interpreting studies share a strong interest in further exploring SA as a useful and powerful tool. Research from both disciplines has already identified some factors moderating SA accuracy, for example, students' cultural backgrounds, previous training experiences and psychological traits, with self-efficacy standing out as an influential variable. To further explore the intricate relations between self-efficacy and SA accuracy, the current study examines the following two hypotheses.

Hypothesis a: Students' interpreting self-efficacy has a significant impact on their SA accuracy in the context of an undergraduate interpreting classroom.

Hypothesis b: Students with higher ISE tend to be more accurate in their SA.

4. STUDY

4.1. Participants

A total of 53 English major students participated in the study, coming from two parallel classes that the researcher taught. They were in their senior year at a regular university in central China. Among them, 45 are female; 8 are male. All were at a similar age of around 22 years old. All completed an introductory course of Chinese/English consecutive interpreting one year earlier. The Advanced Interpreting course, during which the study took place, was a 17-week course and was compulsory despite students' personal interests or linguistic proficiency. Thus, this helps avoid investigating a group of similarly strongly interested and motivated students that might enrol should the course be optional. The course was theme-oriented, requiring students to write their own bilingual speeches based on pre-set themes and to take turns in performing consecutive interpreting. Given the idiosyncratic linguistic competence and personal attitudes toward interpreting, all levels of performance were expected. Absences and lack of attention to the course by students were not uncommon throughout the semester.

4.2. Instruments

Two grids were employed in the current study: one was for measuring participants' interpreting self-efficacy; the other was used to self-assess interpreting performance.

4.2.1. *Interpreting Self-Efficacy Scale (ISES)*

The scale is adapted from Lee's ISE scale targeting South Korean students (Lee, 2014). The scale, comprised of 21 items, consists of three sub-sections: evaluating self-confidence, self-regulatory efficacy and preference for task difficulty, all of which have high Cronbach's alpha values, with the overall coefficient reaching .896. A six-point Likert scale is placed alongside the descriptors, ranging from "very untrue of me", "untrue of me", "somewhat untrue of me", "somewhat true of me", "true of me", to "very true of me", with accompanying scores of 1 to 6. Higher scores on the items mean higher levels of ISE. However, 17 items are negatively worded, which requires reverse scoring.

The original scale was in English, but to facilitate understanding, the researcher provided a standard Chinese translated version alongside the original one, making the final scale bilingual. The participants were advised to refer to the English original should there be any confusion.

Taking into consideration the time and cost of distributing and collecting responses, ISES was administered digitally via an online platform¹. Besides all the items related to ISE, only the participants' names were required because of the need to match their ISE scores with SA scores.

4.2.2. *Self-Assessment Grid (SAG)*

An SA grid for simultaneous interpreting was developed by Schjoldager (1996), later undergoing various modifications for different contexts and research purposes. This grid was presented to students with another SA grid specifically measuring Chinese undergraduate students' competence in consecutive interpreting². It was assumed that students would choose the second one as it fitted the Chinese context. However, the latter one was dismissed due to its highly academic wording and the mismatch between its description and students' actual competence. After several rounds of pilot SA and discussion, Schjoldager's SAG was modified and finalised.

The SAG comprises four sub-sections and 16 items in total, assessing comprehensibility and delivery, language, coherence and plausibility, as well as the loyalty of an interpreter. Students are required to mark themselves on each item from 0 to 5. The higher the marks, the better they think they have done. Instead of focusing on the overall competence of students, this SAG targets detailed performance every time students practice. By marking and noting down examples beside each item, students are given the opportunity to reflect and improve.

4.3. Procedure

After finalising the SAG, students underwent two rounds of SA training, one in English-Chinese consecutive interpreting, the other in Chinese-English. Through a smart learning

¹ The platform, wenjuanxing (wjx.cn, roughly translated into Star of Questionnaires), is a massive, free, and open platform for questionnaire design, distribution and data collection.

² For this grid, please refer to Luo, M. (2018). *Developing a self-evaluating scale of requisite consecutive interpretation competence for interpreting students in China's Universities* [Doctoral dissertation, Shanghai International Studies University].

app³, students downloaded the source text recordings and interpreted multiple times while recording. Then they chose the best version to upload, along with the SAG, after noting down questions and mistakes. The teacher selected some recordings and SAGs as samples to give feedback, make comments and point out common mistakes during classes.

In the final week before the end-of-term exam, all students were required to interpret an English-Chinese segment and self-assess as part of the routine practice. But before they interpreted, they were informed of the topic, basic information and requirements to participate in the current study. Anonymity and confidentiality were guaranteed and students were assured that non-participation would not affect their summative scores. Only those who gave consent proceeded to fill in the ISES by using mobile phones to scan a quick-response code presented on a PowerPoint slide. Then all students moved on to complete the interpreting and SA. All the requirements were similar to those in previous training except that students recorded their interpretation simultaneously without a second chance. After all the recordings and SAGs were uploaded, those from the participants were extracted and analysed. Altogether 55 students out of a total of 60 took part in the study. Two students failed to upload their recordings, making the final data set contain 53 students' ISE scores and SA scores.

4.4. Data analysis

Students' ISE scores were compared to their SA accuracy by their SA scores minus the teacher's scores before employing correlation tests to look for any significant relations between variables. Stata 14, a similar statistics tool to SPSS, was used in this process.

5. RESULTS

A preliminary attempt to correlate ISE with SA accuracy by drawing a scatter plot to delineate any potential relations was undertaken. As Figure 1 shows, the correlation was not immediately clear. Thus, further computation was required.

³ The application, Chaoxingxuetitong (roughly translated into Super Star Master of Learning), provides platforms for teacher-student interaction and smart learning via mobile phone portals and web page portals. Teachers and students can easily access the portals and sign up for virtual classes, thus arranging course notices, up/downloading assignments, checking progress, etc.

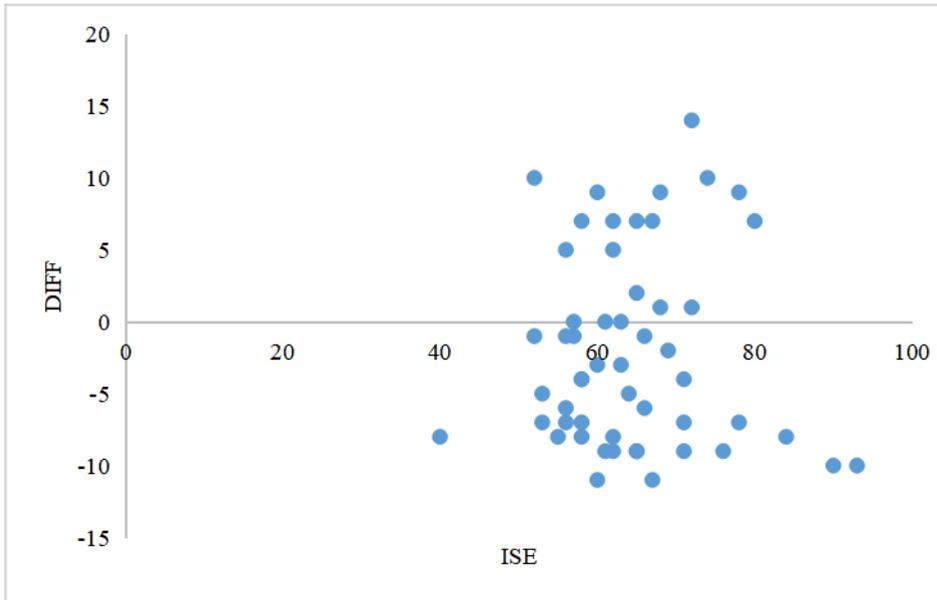


Figure 1. ISE-SA accuracy scatter plot.

Note. DIFF= student mark - teacher mark

Since neither ISE nor SA accuracy was normally distributed, a Spearman correlation test was performed. Table 1 presents the results. Surprisingly, although ISE was negatively correlated with assessment difference⁴, the relationship did not reach a significant level. Further analysis revealed that students’ SA and teacher’s assessment was significantly positive while students’ ISE was not related to their SA, which was also contrary to previous perceptions.

A closer look at the original data yielded some unexpected findings (Table 2). Given the maximum score of the ISES is 126, this pool of students had a medium to low level ISE even though some students produced quite excellent performances. From the table, it is noted that even the highest score (93) barely falls into the third quartile⁵, a far cry from the commonly perceived high ISE level. Similarly, students tended to be stringent in assessing themselves, provided the full mark for the SAG is 80. With 66 being the highest mark, it seems most students were not satisfied with their interpretation. The teacher’s marks presented a similar trend, as SA was correlated with the teacher’s assessment. From this perspective, students in the current study were quite accurate in their SA because the difference between their marks and the teacher’s marks was narrow, yet a convincing correlation was

⁴ As SA accuracy is exhibited by the score difference between SA and teacher’s scores, the larger the assessment difference is, the less accurate the SA is.

⁵ First quartile 0-31; second quartile 32-63; third quartile 64-95; fourth quartile 96-126

not established between students' SA accuracy and their ISE. When the inspection narrowed down to students with relatively "high" ISE levels (scores over 63, see Table 3) the original observation stood. Despite the negative correlation between ISE and SA accuracy, the significance was not confirmed.

Table 1. Spearman correlation (ρ)

(N= 53)	ISE	DIFF	SA	TA	P
ISE	/				
DIFF	-0.02	/			0.91
SA	0.16		/		0.25
TA			0.68	/	0.00

Note. TA = teacher assessment

Table 2. Descriptive statistics

(N= 53)	M	SD	MIN	MAX
ISE	64.42	9.77	40	93
SA	44.21	8.82	24	66
TA	46.23	8.14	28	60

Table 3. Spearman correlation (ρ) for ISE scores over 63

(N= 25)	ISE	DIFF	SA	TA	P
ISE	/				
DIFF	-0.08	/			0.71

6. DISCUSSION

Self-efficacy is widely believed to be able to predict students' academic achievement (Wang et al., 2014). More specifically, in interpreting classrooms, ISE is found to predict students' interpreting performance. When students are asked to assess their own performance, the question boils down to the potential role of ISE in informing their SA behaviour. Thus, the two hypotheses proposed earlier are discussed as follows.

The first hypothesis was rejected. Significant correlations were not found between students' ISE and their SA accuracy in the current case. It cannot be assumed that students' ISE levels were able to predict their SA accuracy. In essence, students' ISE did not mediate how they assessed themselves. The students in this study were mostly accurate in their SA

compared to the teacher's marks. However, most of them were found to be less efficacious than otherwise expected, which was partially reflected in their extremely strict SA marks. According to Table 1, students' ISE levels were only weakly correlated with their absolute SA marks, indicating that the higher a student's ISE, the higher the self-reported SA, but only with a limited amount of certainty. Dominant beliefs hold that students with strong self-efficacy tend to mark themselves higher and exhibit more confidence followed by more efforts and perseverance (Cassidy, 2007; Chan & Lam, 2010), which cannot be proved here.

On the one hand, some scholars argue (e.g., Bolaños-Medina et al., 2018) that self-efficacy is future-oriented, referring to prospective actions and performance, not past performance. But this is exactly the basis for SA. Perhaps the findings here could be illustrated to some extent by the differences in the orientation of ISE and SA. However, Bandura (1994, p. 4) also suggests that self-efficacy can influence a person's SA, but not necessarily in a good way. He points out that people refer to existing knowledge and past experience to test and revise their judgements against short-term and long-term actions while formulating an understanding of the factors tested and their performance. In this process, when under mounting pressure, those with low self-efficacy may suffer from flooding self-doubt that results in irrational analysis and assessment of their own performance. This, in turn, leads to deteriorating performance. And those with high self-efficacy will maintain consistently good analytic thinking, which contributes to better performance and resultant SA. In the current study, the students were believed to suffer from relatively low ISE. Thus, when they were placed under much pressure (i.e., interpreting while recording their own voices), they naturally fell victim to snowballing self-questioning, producing unsatisfactory performance. Therefore, although their SA was close to the teacher assessment, their ISE failed to predict their performance.

On the other hand, multiple factors add to the (in)accuracy of students' SA. To start with, it is observed that interpreting directionality and students' familiarity with the target language will affect their SA accuracy (Han & Riazi, 2018). Chinese students are found to be more accurate when assessing English-to-Chinese interpretation than the other way around because they are more confident and competent in assessing their mother tongue. In the current study, it was assumed at the outset that students would score well on several items in the SAG describing grammar, pronunciation of their Chinese etc. since they are all native speakers. However, quite a few participants under-reported scores on these descriptors, arousing suspicion whether they misunderstood the items or were simply diffident.

Secondly, training and practice are believed to increase students' SA accuracy over time (Chen, 2008; Han & Riazi, 2018; Hung, 2019). The common practice is to ask students to self-assess at least twice before the formal experiment (Falchikov, 2004; Stefani, 1994). The current study was in full compliance with this practice, guiding the participants through the SA of English-Chinese bidirectional interpretation followed by feedback and sample analyses. In the tutorial, the frequent cases of students' under-reporting were pinpointed at an early stage and specifically addressed. The teacher went to great lengths to reassure students that they would interpret well as long as they kept up the good work. Moreover, the items in SAG over which students and the teacher produced the most divergent marks were singled out and discussed in-depth. Yet in the formal experiment, some participants still maintained harsh and critical attitudes towards themselves while some expressed confusion over the wording of certain items. This was despite discussion sessions where students' opinions

and questions were solicited and misunderstandings were clarified. It seemed that not all participants benefited from the SA training sessions and more efforts are required to address these in future research.

Lastly, the cultural background and other psychological variables of the participants need to be factored in. Chinese students tend to be modest and obedient, which has been regarded as a virtue for a long time (Chen, 2008). When reflected in SA and ISE, this might lead to under-reporting and highly stringent self-perception. Based on everyday classroom observation, this stayed true for the participants in this study. Meanwhile, other psychological variables, including anxiety, might also inform students' SA accuracy (MacIntyre et al., 1997; Shaban et al., 2016). The participants were extremely nervous when interpreting, as shown either by direct communication with the teacher or revealed by their performance in the classroom and the final summative examination. If stress and anxiety are inherent, naturally students' ISE and SA would be affected.

The second hypothesis *per se* turned out to be null and void since none of the participants could be categorised into the "high" ISE group and there was no linear nor positive correlation between ISE and SA accuracy. When the data were inspected more closely, SA accuracy with ISE over 63 again failed to be correlated, denying any mediating role of ISE in SA performance. According to previous research, the higher students estimate their marks to be, the less accurate is their SA; and the higher the student's actual mark, the greater is their SA accuracy (Cassidy, 2007, p. 321-322).

Existing research reports a significant correlation between students' self-efficacy and their performance, but with certain prerequisites. Positive correlations could only be established when students' self-efficacy and their competence are strong enough (Lee, 2018). In interpreting classrooms, self-efficacy is found to be relevant in predicting the performance of students with high linguistic competence and not relevant for underachievers. However, students with high linguistic competence lacking self-efficacy do not produce good performance (Ivars et al., 2014). Likewise, Lee (2018) observes that the higher the ISE, the more confidently ISE can predict performance, while warning it never means ISE exerts any direct impact on students' performance or vice versa. This can be partly evidenced in the current study, given the poor link between the median-to-low ISE across the participants and their overall trend of under-reporting. Obviously, ISE cannot be applied to predict their performance and its mediating function cannot be specified.

But it is also contended that there are other divergent factors moderating SA accuracy. Any student group will be characterised by a variation in aptitudes, experience, ability, etc. and it would be unrealistic to expect a uniform profile of these variables and SA competence across any group (Cassidy, 2007, p. 325). Therefore, it would be more rational not to ignore other concurrent factors when investigating one particular variable, which is concurred by many researchers (e.g., Han & Riazi, 2018; Lee, 2018; Liu, 2018).

7. IMPLICATIONS

The implications of the current study should be discussed in light of several limitations. Firstly, the sample size was not large enough to incorporate participants with divergent ISE levels, thus leading to a data analysis marked by lopsided psychological variables. The result

might be more convincing should more students from different backgrounds be recruited. Secondly, SA tutorials were organised only twice, missing out on clarifying in full any misunderstanding or confusion experienced by some students. Thus, for any future research, it is strongly advised to guide participants through multiple rounds of SA training and to ensure that no participant is left in doubt concerning the SA procedure.

Nevertheless, the findings presented above can contribute one piece of the jigsaw to the studies of SA in language classrooms by specifically adopting an empirical approach to align self-efficacy with SA accuracy, which is under explored yet boasting huge potential (Han, 2018; Li, 2018). This study observed that students' ISE did not mediate their SA accuracy in a significant manner. However, students were capable of accurate SA despite low levels of ISE, rejecting the presumption that higher ISE contributed to more accurate SA. It should also be noted that these results need to be examined within an intricate and complex network of learner variables as a plethora of factors, including anxiety, motivation and gender, contribute to learner behaviours (de Bot et al., 2005; Dörnyei, 2006), and their relations are by no means linear or straightforward. The mere fact that ISE leaves a positive but marginal effect on SA accuracy in this study cannot and should not deny the role of ISE altogether. It is essential for students to feel efficacious in interpreting, otherwise they already fail long before the interpretation even starts. This is particularly relevant in undergraduate courses, where most students are far from being perfect bilinguals and therefore may lack confidence in their interpreting abilities (Ivars et al., 2014).

All in all, much remains unknown concerning the moderating role of self-efficacy along with other social and psychological factors in the language learning process, as well as what students with limited self-efficacy should do to maintain optimal learning outcomes. Notwithstanding all these questions, self-assessment proves to be a meaningful pedagogical device that possesses vast potential of navigating the complicated classroom and boosting teaching and learning. Hopefully, more attention and efforts would be directed to this field in the future.

8. REFERENCES

- Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into Practice*, 48(1), 12–19. <https://doi.org/10.1080/00405840802577544>
- Andrade, H., Wang, X., Du, Y., & Akawi, R. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *The Journal of Educational Research*, 102(4), 287–302. <https://doi.org/10.3200/JOER.102.4.287-302>
- Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (Vol. 4, pp. 71–81). Academic Press.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman and Company.
- Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology*, 41(3), 586–598.
- Bolaños-Medina, A., & Núñez, J. (2018). A preliminary scale for assessing translators' self-efficacy. *Across Languages and Cultures*, 19(1), 53–78. <https://doi.org/10.1556/084.2018.19.1.3>

- Bontempo, K., & Napier, J. (2011). Evaluating emotional stability as a predictor of interpreter competence and aptitude for interpreting. *Interpreting, 13*(1), 85-105. <https://doi.org/10.1075/intp.13.1.06bon>
- Boud, D. (1995). *Enhancing learning through self-assessment*. Kogan Page.
- Boud, D., Lawson, R., & Thompson, D. (2013). Does student engagement in self-assessment calibrate their judgement over time? *Assessment & Evaluation in Higher Education, 38*(8), 941-956. <https://doi.org/10.1080/02602938.2013.769198>
- Bourke, R. (2017). Self-assessment to incite learning in higher education: Developing ontological awareness. *Assessment & Evaluation in Higher Education, 1*-13. <https://doi.org/10.1080/02602938.2017.1411881>
- Brown, G., & Harris, L. (2013). Student self-assessment. In J. H. McMillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 367-393). Sage.
- Brown, T., Andrade, H., & Chen, F. (2015). Accuracy in student self-assessment: Directions and cautions for research. *Assessment in Education: Principles, Policy & Practice, 1*-14. <http://dx.doi.org/10.1080/0969594X.2014.996523>
- Cai, R., & Dong, Y. (2015). Interpreter training and students of interpreting in China. *Journal of Translation Studies, 16*(4), 167-191.
- 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. <http://dx.doi.org/10.1080/02602930600896704>
- Chan, J., & Lam, S. (2010). Effects of different evaluative feedback on students' self-efficacy in learning. *Instruction Science, 38*, 37-58. <http://dx.doi.org/10.1007/s11251-008-9077-2>
- Chemers, M., Hu, L., & Garcia, B. (2001). Academic self-efficacy and first-year college student performance and adjustment. *Journal of Educational Psychology, 93*(1), 55-64.
- Chen, Y. (2008). Learning to self-assess oral performance in English: A longitudinal case study. *Language Teaching Research, 12*(2), 235-262. <http://dx.doi.org/10.1037/0022-0663.93.1.55>
- de Bot, K., Wander, L., & Marjolijn, V. (2005). *Second language acquisition: An advanced resource book*. Routledge.
- Dörnyei, Z. (2006). Individual differences in second language acquisition. *AILA Review, 19*, 42-68.
- Dragemark, A. (2006). Learning English for Technical Purposes: The LENTEC Project. In T. S. Roberts (Ed.), *Self, peer, and group assessment in e-learning* (pp. 169-190). Information Science Publishing.
- Falchikov, N. (2004). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. Routledge Falmer.
- Fernández, F., & Zabalbeascoa, P. (2012). Correlating trainees' translating performance with the quality of their metacognitive self-evaluation. *Perspectives: Studies in Translatology, 20*(4), 463-478. <http://dx.doi.org/10.1080/0907676X.2011.629730>
- Han, C., & Riazzi, M. (2018). The accuracy of student self-assessments of English-Chinese bi-directional interpretation: A longitudinal quantitative study. *Assessment & Evaluation in Higher Education, 43*(3), 386-398. <https://doi.org/10.1080/02602938.2017.1353062>
- Huang, S. (2015). Understanding learners' self-assessment and self-feedback on their foreign language speaking performance. *Assessment & Evaluation in Higher Education, 1*-18. <http://dx.doi.org/10.1080/02602938.2015.1042426>
- Hung, Y. (2019). Bridging assessment and achievement: Repeated practice of self-assessment in college English classes in Taiwan. *Assessment & Evaluation in Higher Education, 44*(8), 1191-1208. <https://doi.org/10.1080/02602938.2019.1584783>

- Ivars, A., Catalayud, D., & Forés, M. (2014). Self-efficacy and language proficiency in interpreter trainees. *The Interpreter and Translator Trainer*, 8(2), 167-182. <http://dx.doi.org/10.1080/1750399X.2014.908552>
- Kissling, E., & O' Donnell, M. (2015). Increasing language awareness and self-efficacy of FL students using self-assessment and the ACTFL proficiency guidelines. *Latin American, Latino and Iberian Studies Faculty Publications*, (151), 1-24. <https://scholarship.richmond.edu/lalis-faculty-publications/151>
- Kostons, D., van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction*, 22(2), 121-132. <http://dx.doi.org/10.1016/j.learninstruc.2011.08.004>
- Lee, S. (2014). An interpreting self-efficacy (ISE) scale for undergraduate students majoring in consecutive interpreting: Construction and preliminary validation. *The Interpreter and Translator Trainer*, 8(2), 183-203. <http://dx.doi.org/10.1080/1750399X.2014.929372>
- Lee, S. (2018). Exploring a relationship between students' interpreting self-efficacy and performance: Triangulating data on interpreter performance assessment. *The Interpreter and Translator Trainer*, 12(2), 166-187. <https://doi.org/10.1080/1750399X.2017.1359763>
- Lee, Z. (2015). *The reflection and self-assessment of student interpreters through logbooks: A case study* [Doctoral dissertation, Heriot-Watt University]. ROS Theses Repository. https://www.ros.hw.ac.uk/bitstream/handle/10399/3033/LeeZ_1015_sml.pdf?sequence=1&isAllowed=y
- Li, X. (2018). Self-assessment as 'assessment as learning' in translator and interpreter education: Validity and washback. *The Interpreter and Translator Trainer*, 1-20. <https://doi.org/10.1080/1750399X.2017.1418581>
- Liu, X. (2018). A longitudinal study of dynamic changes in and contributing factors of learner belief of Chinese foreign language learners. *English Language Teaching*, 11(7), 61-70. <https://doi.org/10.5539/elt.v11n7p61>
- MacIntyre, P., Noels, K., & Clement, R. (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. *Language Learning*, 47(2), 265-287.
- McMillan, J., & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons*, 87(1), 40-49. <https://www.jstor.org/stable/42923742>
- Ng, J., & Earl, J. (2008). Accuracy in self-assessment: The role of ability, feedback, self-efficacy and goal orientation. *Australian Journal of Career Development*, 17(3), 39-50. <https://doi.org/10.1177/103841620801700307>
- Orsmond, P., Merry, S., & Reiling, K. (1997). A study in self-assessment: Tutor and students' perception of performance criteria. *Assessment & Evaluation in Higher Education*, 22(4), 357-369. <http://dx.doi.org/10.1080/0260293970220401>
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66(4), 543-578.
- Pan, J., & Yan, X. (2012). Learner variables and problems perceived by students: An investigation of a college interpreting programme in China. *Perspectives: Studies in Translatology*, 20(2), 199-218. <http://dx.doi.org/10.1080/0907676X.2011.590594>
- Panadero, E., Brown, G., & Strijbos, J. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, 28(4), 803-830. <http://dx.doi.org/10.1007/s10648-015-9350-2>
- Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, 22, 74-98. <http://dx.doi.org/10.1016/j.edurev.2017.08.004>

- Panadero, E., Tapia, J., & Huertas, J. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences, 22*, 806-813. <http://dx.doi.org/10.1016/j.lindif.2012.04.007>
- Panadero, E., Tapia, J., & Reche, E. (2013). Rubrics vs. self-assessment scripts effect on self-regulation, performance and self-efficacy in pre-service teachers. *Studies in Educational Evaluation, 39*, 125-132. <http://dx.doi.org/10.1016/j.stueduc.2013.04.001>
- Paulhus, D. (1991). Measurement and control of response bias. In J. Robinson, P. Shaver, & L. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). Academic Press.
- Postigo Pinazo, E. (2008). Self-Assessment in Teaching Interpreting. *Traduction, terminologie, rédaction, 21*(1), 173-209. <https://doi.org/10.7202/029690ar>
- Raaijmakers, S., Baars, M., Paas, F., van Merriënboer, J., & van Gog, T. (2019). Effects of self-assessment feedback on self-assessment and task-selection accuracy. *Metacognition and Learning, 14*, 21-42. <https://doi.org/10.1007/s11409-019-09189-5>
- Reybroeck, M., Penneman, J., Vidick, C., & Galand, B. (2017). Progressive treatment and self-assessment: effects on students' automatization of grammatical spelling and self-efficacy beliefs. *Reading and Writing, 30*, 1965-1985. <https://doi.org/10.1007/s11145-017-9761-1>
- Schjoldager, A. (1996). Assessment of simultaneous interpreting. In C. Dollerup, & V. Appel (Eds.), *Teaching translation and interpreting: New horizons* (pp. 187-195). John Benjamins Publishing Company.
- Shaban, S., Aburawi, E., Elzubeir, K., Elango, S., & El-Zubeir, M. (2016). Factors influencing medical students' self-assessment of examination performance accuracy: A United Arab Emirates study. *Education for Health, 29*, 75-81. <http://dx.doi.org/10.4103/1357-6283.188688>
- Stauffer, L. (2011). ASL students' ability to self assess ASL competency. *Journal of Interpretation, 21*(1), 80-95. <http://digitalcommons.unf.edu/joi/vol21/iss1/7>
- Stefani, L. (1994). Peer, self and tutor assessment: Relative reliabilities. *Studies in Higher Education, 19*(1), 69-75. <http://dx.doi.org/10.1080/03075079412331382153>
- Wang, C., Kim, D., Bai, R., & Hu, J. (2014). Psychometric properties of a self-efficacy scale for English language learners in China. *System, 44*, 24-33. <http://dx.doi.org/10.1016/j.system.2014.01.015>
- Yan, X., Pan, J., & Wang, H. (2018). The Interplay between foreign language anxiety, willingness to communicate and other learner factors in tertiary interpreting classrooms. In X, Yan, J. Pan, & H. Wang (Eds.), *Research on translator and interpreter training: A collective volume of bibliometric reviews and empirical studies on learners* (pp. 147-168). Springer. <https://doi.org/10.1007/978-981-10-6958-1>
- Yan, X., & Wang, H. (2012). Second language writing anxiety and translation: Performance in a Hong Kong Tertiary Translation Class. *The Interpreter and Translator Trainer, 6*(2), 171-194. <https://doi.org/10.1080/13556509.2012.10798835>

9. ACKNOWLEDGEMENTS

The author would like to thank the two anonymous reviewers and the editors for providing insightful comments on earlier drafts of this paper. Special thanks also go to all the students who willingly participated in this study.