

## Article

# Sustaining Learners' Writing Development: Effects of Using Self-Assessment on Their Foreign Language Writing Performance and Rating Accuracy

Xiaoyu Sophia Zhang <sup>1</sup>  and Lawrence Jun Zhang <sup>2,3,\*</sup> 

<sup>1</sup> The Mind Lab, Auckland 1023, New Zealand

<sup>2</sup> School of Foreign Studies, Xi'an Jiaotong University, Xi'an 710049, China

<sup>3</sup> Faculty of Education and Social Work, University of Auckland, Auckland 1023, New Zealand

\* Correspondence: lj.zhang@auckland.ac.nz; Tel.: +64-09-623-8899

**Abstract:** Although the benefits of using self-assessment on student writing performance have received wide recognition over the past two decades, minimal research is available on the effects of using self-assessment of writing on English as a foreign language (EFL) students' writing performance, especially in the tertiary context of China, where such research is still in its infancy. To fill the abovementioned lacuna, the present study adopted a quasi-experimental approach to implementing a self-assessment-based intervention in Chinese tertiary EFL writing classes. Specifically, students were randomly assigned into either an intervention class that implemented self-assessment or a comparison class that employed peer assessment as classroom practice to promote students' sustainable development of writing skills. Quantitative and qualitative data were collected, and the research findings indicated that the intervention group experienced a larger increase in their holistic writing performance and rating accuracy when compared with the comparison group. Furthermore, the qualitative findings reveal students' enhanced rating accuracy after the intervention. This study contributes to research on self-assessment in the EFL writing domain as a basis for further deliberation on self-assessment in higher education, and it also provides much needed empirical evidence for the potential value of student-centred sustainable assessment approaches such as self-assessment. Findings also provide teachers with pedagogical implications for developing sustainable and capable self-assessors of writing.

**Keywords:** self-assessment; EFL writing; writing performance; rating accuracy; sustainable development of writing skills



**Citation:** Zhang, X.S.; Zhang, L.J. Sustaining Learners' Writing Development: Effects of Using Self-Assessment on Their Foreign Language Writing Performance and Rating Accuracy. *Sustainability* **2022**, *14*, 14686. <https://doi.org/10.3390/su142214686>

Academic Editor: Antonio P. Gutierrez de Blume

Received: 1 October 2022

Accepted: 1 November 2022

Published: 8 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Over the last two decades, there has been a growing recognition that widening students' engagement in the learning and assessment process is a vital contributor to their future success [1–4]. Internationally, therefore, student-centred assessment forms, such as self-assessment and peer-assessment, are receiving more attention than before despite their challenging and time-consuming nature [5–8]. It is argued that a prerequisite for students to improve their learning is to develop the capacity to supervise the quality of their learning [8]. Likewise, as Boud [5] contended persuasively, “the ability to self-assess is a key foundation to a career as a lifelong learner who can continue their education after formal education has ended” (p. 14). Further, the ability to self-assess is critical for students to become efficient self-assessors of their learning and sustain sophisticated levels of self-assessment preparing for unfamiliar professional situations in the future [9–12].

In educational research, self-assessment is considered as a salient tool to empower learners to organise, monitor, and sustain their learning; the positive effects that self-assessment may have on learning and academic achievement are recognised, and research interest continues to grow in the field [3,13–15]. The advantages of self-assessment have

already been acknowledged in a range of disciplines in recent years, and self-assessment is considered as a key strategy to develop and sustain students' self-assessment capabilities and self-regulated learning skills [16–18], especially in different writing stages to heighten students' awareness of the ownership of their writing [10,19–21]. It is posited that self-assessment brings a range of benefits to students' learning. Firstly, self-assessment facilitates students to achieve better test performance, a deeper understanding of their work, and a higher task quality. Secondly, self-assessment fosters a more responsible and reflective learner, as compared with peer or teacher assessment [22]. Some researchers have claimed that peer assessment and teacher assessment as two significant forms can be used together to serve as catalysts for students' self-assessment for improving learning [2,5,21] by “clarifying where students' misunderstanding lie” [23] (p. 320). Thirdly, self-assessment affords students an opportunity to obtain power or control over their learning [24] through learning to assess and assessing to learn, in contrast with learning contexts in which they have no power [25]. In doing so, students take responsibility for their learning and are committed to improving themselves as learners [7,26–28].

False self-assessment could possibly lead to students' inaccurate understandings of their abilities, learning outcomes, and curricular expectations in the assessment process, and, over the last three decades, there has been increasing research investigating the role of self-assessment in writing at different learning levels, as self-assessment and writing performance are deemed to be as closely associated with one another [5,10,21]. Nevertheless, it is surprising to find that research on the effects of using self-assessment in the EFL contexts has not been sufficiently documented, especially in the domain of tertiary writing [29]. For instance, in western contexts, most research about student self-assessment focuses on the elementary or the secondary level [11], and empirical research on self-assessment conducted in tertiary EFL writing classrooms has received limited attention. Furthermore, even though students' rating accuracy in self-assessment of writing in relation to that of teacher's rating has been a major area of interest in previous studies, it has rarely been examined from a qualitative lens; specifically, the similarity of students' and raters' comments on the same piece of written work regarding students' writing strengths and weaknesses has been neglected. Given that research gap, this research aims to extend the understanding of one important aspect of foreign language learning, namely, the use of self-assessment to sustain EFL writing development. Considering that studies on student self-assessment of writing are relatively scarce, a quasi-experimental approach was used to explore students' writing development through the lenses of self-regulated learning theory and formative assessment theory [30]. It is expected that empirical evidence from the self-assessment of writing practices would increase teachers' willingness to implement self-assessment, deepen students' understanding of their writing quality, and contribute to not only sustainable writing programmes, as well as other aspects of learning ultimately.

Undertaken in a medium size Chinese university, 92 English major sophomore students and two lecturers participated in the current study. Findings of this research are a possible starting point to provide much needed empirical evidence of the value of student-centred formative assessment and to understand how self-assessment of writing practices could influence and sustain learners' writing performance and rating accuracy in an EFL context.

## 2. Literature Review

### 2.1. Defining Self-Assessment in the Writing Context

It appears that there is no uniform definition of standard self-assessment due to its multifaceted nature, as well as varied forms of its implementation, such as self-grading, self-revision, self-reflection, and self-feedback [13,31–33]. Over the last three decades, although widely varying definitions have emerged, self-assessment is generally referred to as a “variety of mechanisms and techniques through which students describe (i.e., assess) and possibly assign merit or worth to (i.e., evaluate) the qualities of their own learning processes and products” [32] (p. 804). Self-assessment has also been categorised into performance-

oriented and development-oriented assessments [34]. Echoing Oscarson's [34] proposal, Tan [35] further classified self-assessment into three types (teacher-driven, programme-driven, and future-driven), according to teachers' and students' power distribution, as well as the extent of students' contribution during the self-assessment process. All three types of self-assessment could lead to the development of student self-assessment skills; however, future-driven student self-assessment is deemed as more important than others because it emphasises students' motivational development and the sustainable development of their self-assessment capacity beyond the limits of the course content and requirements [9].

In the context of foreign language writing, which is an important means for learners to acquire foreign language skills [33,36,37], self-assessment has been promoted as a valuable self-regulated learning strategy [15–17], and it can be further interpreted from two perspectives, namely, formative and summative. The formative perspective tends to focus on the learning processes, whereas the summative view emphasises on learning outcomes [38,39]. Regarding the formative uses of self-assessment, for example, students use self-assessment to generate feedback for themselves before formal grading, which is promoted as a valuable form of sustainable assessment [6,30,40,41]. Self-assessment, in that sense, refers to a self-regulatory and cyclical process in which learners create and experience evaluation concurrently with minimal reliance on teachers' support. Specifically, students are responsible for collating and reflecting on information about their knowledge, performance, and accomplishment while learning [33]. Then, prior to formal evaluation [42], students identify and evaluate the possible approaches to improving different learning aspects against individual or pre-established criteria or rubrics to revise their work accordingly [22,43–45]. In this process, some scholars have argued that self-scoring or grading one's own work should not be involved as those practices implied summative self-assessment [32,40,46,47], which may pressure students to make a hasty judgement on either work quality or task proficiency and, therefore, overlook the importance of sustainable learning development [5,48].

Nevertheless, summative self-assessment practices, if applied well by language teachers in writing classrooms, can be effective measurement tools not only to supplement formative self-assessment but also to empower and encourage students to reflect on and improve their performance [30,49,50]. For example, some researchers have argued that, before making more accurate self-judgements, students need to develop their self-assessment competence progressively and that simple self-scoring or grading practices (summative self-assessment) could be a useful starting point to sustain realistic and sophisticated self-assessment in the future [51].

## 2.2. Teachers' Role in Student Self-Assessment

To promote students' agency in the process of self-assessment, the complexity of the teachers' role needs to be clarified from several aspects. Firstly, teachers need to relinquish control over students and adopt a mediating and modelling role in the assessment process. At the same time, they also need to understand that sharing and distributing their power to students during self-assessment is beneficial [32,52]. Secondly, in addition to imparting relevant self-assessment knowledge to students [21], teachers' roles are not limited to counsellor, materials developer, administrator, and organiser in the classroom [19,53]. For instance, teachers should also sustain students' dignity or self-efficacy when they turn to teachers for feedback regarding their self-assessment to ensure students that their teachers are confident in their ability to self-assess their own writing, so that their performance in relevant tasks can be gradually enhanced [33,54]. Nevertheless, "a fine-tuned self-assessment ability does not come automatically to all students" [55] (p. 675), and more teacher intervention is expected in early stages to help students internalise self-assessment. Hence, in addition to the abovementioned roles, a critical but challenging part of teachers' roles is how to integrate self-assessment systematically and effectively in the classroom by scaffolding students in self-assessment procedures.

In the writing context, successful teachers' scaffolding, such as awareness raising, timely guidance, role modelling, and co-construction of self-assessment criteria, is the key

for students' self-assessment development [16,19,56,57]. Pivotal to the effective implementation of scaffolding is for teachers to provide students with an explicit explanation of why, how, and when to self-assess, followed by relevant activities and well-timed feedback to support students' actualisation of realistic self-assessment [30].

As for implementing self-assessment, similar three-level models were proposed [46,58] for teachers with increasingly less teacher involvement in each level. For instance, as shown in Table 1, students are engaged in four stages of self-assessment with the teacher transferring, gradually, more responsibility and freedom to students, to enable students to focus on the quality of their work rather than a grade [58]. It can be seen that, across all levels of self-assessment, teachers' feedback is equally important, particularly the timing and the involvement of encouragement. This is because students need sufficient time to digest teachers' feedback for further actions and external assurance to enhance their confidence in practising self-assessment [25,30,32]. For example, at the beginning level, only with detailed and transparent teacher feedback can students make progress in their learning and understand how to interpret feedback, make connections between the feedback they receive and the work they produce, and then apply the feedback to improve their future work [21,47]. Providing students with personalised, diagnostic, and encouraging feedback can be challenging for teachers [19], especially when it comes to using feedback to help students overcome their instincts of ego protection and avoid possible ego inflation, resulting in students' overestimation or underestimation during self-assessment [7,28]. In other contexts such as the EFL settings in China, a major challenge for educators to support student self-assessment is that English is a foreign language, whose script, syllables, and grammar could be different and difficult for them, not to mention the challenge of performing self-assessment per se in the foreign language [59–61].

**Table 1.** Teacher's role in implementing self-assessment in four stages.

	Stage 1	Stage 2	Stage 3	Stage 4
Levels of implementation	Raising awareness and establishing criteria	Teaching students how to apply criteria	Providing feedback to students on application of criteria	Setting learning goals and strategies
Beginning	Criteria are given to students for their reaction and discussion	Examples of applying criteria given to students	Teacher provides feedback	Goals and strategies determined by the teacher
Intermediate	Students select criteria from a menu of possibilities	The teacher describes and models how to apply criteria	Feedback is provided by both teachers and students	A menu of goals and strategies is provided by the teacher
Advanced	Students generate criteria on their own or with the teacher	Students apply criteria to their own work	Students initiate and justify their own feedback with the teacher's help	The student constructs goals and strategies to achieve the goals with the teacher's guidance

Note. Adapted from [58].

Specific approaches that teachers can employ to scaffold students' practising of self-assessment better and the time that teachers no longer need to provide feedback to support student self-assessment are still not clear [62]. For EFL practitioners, there is a challenge in achieving a balance among creating a congenial and supportive learning atmosphere that frees students from their long-standing disempowerment in the traditional assessment activities [24,52,63], affording students sufficient independence to experience a sense of control, and offering them sufficient guidance and facilitation in self-assessment of writing procedures to achieve their goals [64]. The goal, however, is to ensure that teachers are well supported to help students develop the knowledge and capabilities needed for effective engagement in the self-assessment process.

### 2.3. Accuracy of Student Self-Assessment

In the past two decades, to better understand why there is a tendency that students either underestimate or overestimate themselves during self-assessment [65–67], researchers have conducted a wide range of studies to examine the accuracy, i.e., the validity, consis-

tency, or reliability [13,30,57,68], of student self-assessment in EFL writing. It is surprising to find that the previous literature tended to evaluate the effectiveness of student self-assessment through conformity with teachers' marking rather than the formative and sustainable learning growth that students have displayed in the self-assessment process.

In addition to the fact that using teacher and external judgements as the standard to evaluate students' self-assessment accuracy can be problematic given their imbalanced experience and knowledge bases for assessment [16], multiple reasons could lead to the arguable misalignment between students' and teachers' judgements. Firstly, the subjective nature of self-assessment as a measurement tool has been often questioned for students to use it as novice assessors with limited knowledge and information sources [30,50]. Secondly, social and psychological factors, and cultural background have been shown to affect students' self-assessment accuracy significantly [7]. For instance, students' varied academic ability and linguistic proficiency may prevent them from assessing their foreign language writing effectively [66–74], and some nationalities tend to overrate their language proficiency, whereas others are inclined to undervalue it [75]. To be specific, in Asian countries, being modest about one's ability in the public is highly valued [25,69,76], and giving oneself a good assessment may be interpreted as boasting [70]. Therefore, such cultural values, for instance, modesty and self-confidence, may impede self-assessment in certain contexts and constrain students being honest in self-assessment and, whether self-assessment is performed privately or publicly may also impact the accuracy of students' self-assessment [30]. Thirdly, if only applying the self-assessment outcomes in a summative manner [51,67], such as when using students' self-marking mainly to substitute for teacher marking [18], students' formative development during the sustainable assessment process can be overlooked; therefore, the effectiveness of student self-assessment could be compromised [25]. Fourthly, teachers' reluctance to promote self-assessment because of their lack of assessment literacy and limited experience in using effective student-centred practices [19,77] may result in superficial self-assessment implementation and, therefore, not lead to students' actual improvement [32,59].

The above reasons might explain why inconsistent findings were often reported in the literature concerning the accuracy of students' self-assessment when compared with teacher or external assessment [34,65,67,70]. For instance, in some studies, a lack of congruence was reported between students' self-assessment scores and those of the teachers, and students' self-assessment marks were not found to be comparable to their actual academic abilities [8,71]. Contrary to the literature that suggests a lack of alignment between student self-assessment and teacher assessment, other studies reported relatively good levels of agreement between students' and teachers' grading [12,67,72,73]. For example, in Leach's study [67], when university students ( $n = 472$ ) were provided with the opportunity to self-assess their work, the results showed highly significant statistical correlations between students' and the teacher's rating.

It is, therefore, argued that the effectiveness of self-assessment should not be judged by the conformity between students' and teachers' marking, and academics may need to place more emphasis on students' self-assessment reflexivity and sustainability. For example, the focus should be on how EFL students have developed during the assessment process in their ability to identify their strengths and weaknesses [25,72], rather than being concerned about its accuracy [35]. Therefore, greater attention should be paid to the learning function of self-assessment [78] regarding how students learn to self-assess and how teachers support students' self-assessment rather than comparing students' self-generated grade with the external assessment [79,80]. The above factors should only serve as aspects to consider in students' self-assessment practices, rather than the basis on which to reject self-assessment [81] because the positive role self-assessment can play in learning outweighs this possible lack of accuracy.



#### 2.4. Use of Self-Assessment in the Writing Context

Arguably, self-assessment and writing are closely connected in a range of ways. For example, writing features an ideal platform for students to practise self-assessment by analysing and evaluating their written work [82] through which students are more likely to develop a proactive and critical stance in their writing [83,84]. Over the last three decades, a substantial body of research has also discussed how self-assessment could be used in the writing context at different learning levels [19]. However, such a research area is still considered to be in its infancy with many unknown subfields [10,77]. For example, although previous studies employing self-assessment in the writing class have shown the positive effects of self-assessment on students' writing performance, few studies have implemented systematic self-assessment in the Chinese tertiary writing classrooms, and English-major students are rarely involved [63,85].

Among previous studies, three major areas could be identified concerning the application of self-assessment in writing, namely, students' self-assessing accuracy in relation to that of the teacher's assessment, the effect of using self-assessment practices on students' overall learning achievement, and students' self-assessment of their writing against certain criteria or rubrics, with the last one seeming to be the most prevalent practice adopted [84,85].

In the L1 contexts, the US, the UK, New Zealand, and Australia seem to lead most studies on self-assessment of writing over the past two decades (see [69,86] for reviews). For instance, positive correlations were found when comparing tertiary students' and lecturers' overall rating grades on students' assignments with the same marking guide used, even though students were not properly trained prior to the assessment [12,23]. In contrast, self-assessment of writing is usually performed with training, especially with learners at lower school levels. Ross et al.'s [11] quasi-experimental study is an example that helps illustrate our point. In their study, fourth–sixth-grade students ( $n = 296$ ) were divided into a treatment group and a control group. Students in the treatment group were taught how to self-assess their narrative writing against a predetermined rubric (60 min/day) for 8 weeks, whereas the control group was not. After 8 weeks, it was indicated that, even though weak writers in treatment groups only slightly outperformed the control groups in narrative writing performance (effect size = 0.18), they had become significantly more accurate in self-assessment. The small impact might be because the students were using a different rubric to grade their work and that they were not familiar with that rubric. In a later study, Andrade and Du [46] provided a relatively in-depth understanding of undergraduate students' experiences of using rubric-referenced self-assessment from a qualitative perspective. Students who were interviewed reported the positive impact that self-assessment had on their work quality. In a recent study, students' experiences and understanding on the effectiveness of self- and peer-assessment in a geography course were investigated through surveys [87]. Similar to what was found in Andrade and Du's [46] study, the findings also showed that students were mostly favourable to the benefits concerning quality that practicing self-assessment brought to their writing. However, students also criticised the arduous and time-consuming nature of self-assessment.

In terms of the EFL context, it appears that little attention has been paid regarding how self-assessment might contribute to the development of EFL writing in early research, but more EFL researchers have started considering the application of self-assessment in the writing domain, especially in Asian and African countries [78,88,89]. In addition to examining the agreement between students' and teachers' grading, a range of studies have explored students' and teachers' perceptions and experiences during self-assessment of writing [59] to not only further affirm the benefits of using self-assessment in the writing classroom, but also identify the possible reasons for teachers' and students' hesitance or resistance to self-assessment [29,85]. A series of Asian studies have also investigated the effects of applying self-assessment in various forms, such as self-assessment used alone [50,72], self-assessment paired with peer assessment [90], and self-assessment used with both peer and teacher assessment [76,88]. Some of these studies [76,90] reported con-

siderable benefits from implementing self-assessment in the EFL writing class; however, the findings were not helpful in identifying the effects of implementing self-assessment alone on students' writing performance, as self-assessment was only part of a multipronged intervention. Mazloomi and Khabiri's [72] study filled that gap by using a quasi-experimental approach with only self-assessment as an intervention to improve students' writing skills, and their findings concurred with those of Mok et al. [91], who indicated that continuous students training and teachers feedback during self-assessment are the prerequisites for writing improvement.

Similar to the L1 and other EFL contexts, researchers in the Greater China region have also empirically examined the use of self-assessment in the writing domain. For instance, in Hong Kong, where the education system and learning assessment culture are somewhat different from that in mainland China due to the influence of British colonial rule and its educational traditions [92], self-assessment was mostly performed with young learners in the classroom in limited forms [53,93], but those case studies demonstrated positive effects of implementing such student-centred assessment approaches. Likewise, in mainland China, empirical evidence of young learners' self-assessment trajectories can be found in a number of studies [89,94]. Even though their study did not explain in any detail the processes that young learners used to self-assess their writing, it pointed out that Chinese young learners could self-assess their English reading/writing skills and knowledge truthfully because the correlations between self-assessment of reading/writing scores and reading comprehension/writing test scores were found to be significant.

Only a limited number of studies have focused on university-level students in mainland China [85,95], and Liu [85] appears to be the first study using self- and peer assessment in the writing class, investigating the reliability and validity of self- and peer assessment with second-year university students ( $n = 120$ ). In that study, Liu [85] simply considered English major students with high English proficiency and Japanese major students with low English proficiency, which seemed problematic when explaining why Japanese major students tended to overestimate their writing, whereas English students could self-assess themselves objectively. Some later studies also demonstrated the effectiveness of rubric-referenced self-assessment in tertiary writing classrooms with non-English major students within different timeframes [15,85,95,96], and those findings are useful to gauge Chinese students' responses to varied perspectives of self-assessment of writing. Unlike non-English major students, English major students enrol in English courses specially designed to develop their knowledge of English, especially their competence in using English in various skill areas such as listening, reading, writing, and speaking, as well as a command of the theoretical and content knowledge of linguistics and literary canon (i.e., literature in English) [64].

To promote sustainable writing development, the current study was, therefore, undertaken to address the aforementioned gap by investigating the effects of using self-assessment of writing on EFL tertiary English major learners' writing performance, as well as students' rating accuracy from both quantitative and qualitative lenses. Specifically, the following research questions directed this study ("students" refer to Chinese undergraduate English-major students).

What effects does the use of self-assessment in the EFL writing class have on (a) students' writing performance in different dimensions, and (b) students' rating accuracy in self-assessment of writing as related to the raters' assessment?

### 3. Methods

Contextualised in teacher-centred and exam-oriented English teaching practices in mainland China [97,98], the current study adopts a quasi-experimental approach to investigate the effects of using self-assessment on student participants' writing performance, as well as on the development of students' rating accuracy. Collecting both quantitative and qualitative data, this research adopted a mixed-methods approach (see Table 2 for research design details).

**Table 2.** An overview of the research design.

	Research Aims	Instruments	Participants
Preparatory (piloting) stage	Instrument revision and validation	Timed writing tests and self-assessment tasks using self-assessment of writing rubric	English major students Year 2 ( $N = 10$ )
Main study	Implementation of self-assessment-based writing intervention and examination of its effects on students' writing performance and rating accuracy	Pre- and post-writing tests and ore- and post-self-assessment tasks using self-assessment of writing rubric	English major students Year 2 Intervention group (self-assessment; $N = 51$ ) Comparison group (peer-assessment; $N = 41$ ) English major lecturers Year 2 ( $N = 2$ )

### 3.1. Context and Participants

The current study embedded self-assessment- and peer assessment-based writing instructions in the traditional writing course syllabus of Chinese tertiary EFL writing classrooms over a time span of 4 months. Using convenience sampling, 92 English-major participants (15 male students—16%, and 77 female students—84%) and two teachers volunteered to participate in the current study with their names anonymised. They came from a medium-scaled (around 20,000 students) Chinese university, in which EFL writing teachers spend considerable time marking students' written work. However, unfortunately, such marking and commenting leads to very few improvements in students' subsequent work. It appears that English-major students, even after studying linguistic constructions and different genres in the writing course for a few years, still remain unclear about where they are and how well they are achieving relative to expectations.

At the time of this research, the 92 English major students, shared comparable characteristics, such as similar years of English learning experience ( $M = 9.18$ ,  $SD = 1.24$ ), age ( $M = 19.42$ ,  $SD = 1.15$ ), no overseas learning experience, and previous self-assessment of writing experience. They had just embarked on their second year of study, and they were all enrolled in a compulsory English writing course (conducted once a week; 1.5 h per week for 16 weeks), which focused on linguistic knowledge and various types of English writing. Given the above shared background, homogeneity among student participants was mostly ensured prior to the research; therefore, the potential risk of convenience sampling (at university level) not representing the population was mitigated. Students from the four intact classes were randomly divided into two groups (two classes formed a group) and assigned to either the intervention group (self-assessment) or the comparison group (peer assessment).

As for the two lecturer participants, they were experienced female English language lecturers of a similar age (age = 38 and 39), who taught year two in the selected university. Each of them had a master's degree in English language and/or English literature. With more than 10 years of tertiary teaching experiences, they were confident in ensuring the smooth conduct of the intervention even though they had no previous experience in using either self-assessment or peer-assessment in their writing classes. The two lecturer participants were randomly allocated to the intervention group or the comparison group classes.

### 3.2. Data Collection

A week before the 16-week self-assessment-based teaching intervention, we piloted the writing test (the same topic as would be used in the pre-test and post-test) and the self-assessment rubric with 10 English-major sophomore students (nonparticipants) from the research site to detect, if any, and resolve students' potential difficulties in understanding



the instruments. The instruments were then revised on the basis of students' feedback to ensure they had an accurate understanding of the content.

When the new academic semester started, quantitative data were gathered from the pre- and post-writing tests using the self-assessment writing rubric, and qualitative data were collected from students' and raters' comments on their writing strengths and weaknesses. To explore students' writing performance before and after the intervention, students were asked to complete two given-topic cause and effect essays, writing at least 200 words in response to the prompt (e.g., topic information outline) in 45 min. The cause-and-effect type of essay is considered as a common writing task for Chinese undergraduates [97]. Stemmed from previous examination papers in the Chinese National Test for English Majors, the general test difficulty of the pre- and post-writing topics match the difficulty level and relevance to students' daily lives (see Appendix A for writing topic details). Participants were required to complete the writing tasks independently in class with a pen and paper without any external help; they were at liberty to add, to delete, or to cross out their content, rather than rewrite or erase.

Upon the completion of the writing task, students from both the intervention and the comparison groups were provided with a four-dimension and five-scale self-assessment writing rubric (see Appendix B for details). The four dimensions included task achievement, coherence and cohesion, language resources, and mechanics, which were adapted from previous analytical scoring schemes [99–101] taking into consideration factors that Chinese EFL students perceived as affecting the rubric's effectiveness during their self-assessment of writing [85].

Students were required to self-assess their cause-and-effect essays according to the above dimensions in five scales, generating an individual section score and an overall score according to the provided rubric. The last part of the self-assessment writing rubric asked students to indicate a writing rubric dimension and comment on why that dimension could be a strength or weakness in their writing. The self-assessment task was completed in 20 min. Raters used the same rubric to grade students' pre- and post-writing texts, as well as identifying and commenting on students' strengths and weaknesses.

In addition to investigating the statistical correlation between students' self-assessment and raters' assessment, this study examined students' formative development in rating accuracy by comparing six randomly selected students' self-comments and raters' comments regarding their strengths and weaknesses for possible overlaps or divergence.

Given the limited attention to how self-assessment functions in EFL writing settings, this study implemented and evaluated a self-assessment-based intervention in two contrasting conditions: the intervention group (self-assessment) and a comparison group (peer-assessment). Implementation fidelity of the 4-month writing intervention was ensured from three perspectives. Firstly, both groups utilised the same textbook and writing tasks and received the same class instruction time and teacher feedback during the intervention. Secondly, the teaching delivery quality was ensured by teacher training (through an 8 h workshop) prior to the intervention to upgrade the EFL writing lecturer's knowledge base of self-assessment in EFL writing, and lesson plans were developed at different timepoints featuring both summative and formative uses of self-assessment. Thirdly, we made sure that the intervention group and comparison group teachers did not share their teaching practices, lesson plans, and other related materials during the time when the study was ongoing by asking them to sign an agreement (participant information sheet form) in case they were influenced by each other. Verbal validation from teachers was also achieved at different timepoints to ensure the intervention group applied self-assessment while the comparison group did not.

During the intervention, students in both groups were required to perform self- or peer assessment on their assigned in- and after-class writing tasks every week. Students had opportunities to use the feedback from self- or peer assessment to improve their writing before the work was formally evaluated. Students' self- or peer assessment was not included in their final grade. Furthermore, teachers agreed not to look at students' in-class

self- or peer assessment, because keeping the act of self- or peer assessment partially private to students may promote willing and honest assessment practices [16,30].

### 3.3. Data Analysis

Data in the current study were analysed quantitatively and qualitatively. Rater training was conducted prior to scoring the pre- and post-tests to obtain an accurate estimation of students' writing performance. The two raters, who were experienced Chinese university English lecturers, were blind to the research conditions and unfamiliar with the research design, which ensured the possibility of avoiding potential bias in essay evaluation. Firstly, inter-rater reliability was checked in the trial scoring, with a random sample of 40 students' writing texts distributed to two raters, forming approximately 20% of the entire dataset. Then, the two raters compared their scores and discussed any discrepancies in their evaluation in respect to the scoring criteria. The two raters only started to score the remainder of the writing samples independently when their inter-rater reliability coefficients reached a satisfactory level ( $r = 0.89$ ,  $p < 0.001$ ), which ensured the reliability of the rating outcomes [102].

Data generated from the pre- and post-questionnaires, writing tests, and students' self-assessment tasks were screened and cleaned to ensure response validity before subjecting the data to inferential statistical analyses (e.g., independent and paired-samples  $t$ -tests, and Pearson correlation). Writing samples from the writing tests were transcribed into word processor files before scoring. By doing so, the essay format consistency was ensured, and possible rater bias with regard to handwriting was minimised [103].

The writing test data, together with students' self-assessment of writing outcomes, were entered into a database in SPSS 25, a statistical analysis software programme for Windows. Descriptive analysis (e.g., normality, mean score, and standard deviation) and inferential analyses, such as independent-samples  $t$ -tests, paired-samples  $t$ -tests, and Pearson correlation coefficient ( $r$ ), were checked to identify any discrepancies in students' writing performance, as well as students' and raters' assessment results before and after the intervention (between and within the intervention and comparison groups).

To control the effect that multiple  $t$ -tests have on the error rate, 0.05 was not used as the critical level of significance. Instead, the Bonferroni value was adjusted by the number of  $t$ -tests conducted [104]. In addition, to multiply 0.05 by the number of different comparisons (on the same dataset), the benchmark value of Cohen's  $d$  was adopted to interpret the effect size (small = 0.2; medium = 0.5; large = 0.8) of the independent variable on dependent variables [105,106].

The magnitude of the two variables' linear association was measured using the Pearson correlation coefficient ( $r$ ) [104], which was applied to investigate how similar the scores of students and the raters were in students' written texts in the pre- and post-tests. The effect sizes of the Pearson correlation coefficient ( $r$ ) were small = 0.1, medium = 0.3, or large = 0.5 [105]. As explained earlier, two raters were recruited for essay scoring, and the Pearson product-moment correlation coefficient test was used to calculate the inter-rater reliability of the two raters for this task. The Pearson product-moment correlation coefficient was also applied to check intra-rater reliability for data coding. High correlations represent high reliability.

Potential differences concerning social factors, writing performance, and students' rating accuracy in self-assessment of writing between the two groups were evaluated before the pre-test, using independent-samples  $t$ -tests. Assumptions of  $t$ -tests were satisfactorily met. To protect against type I errors, the Bonferroni procedure was applied to each  $t$ -test with the adjusted alpha (0.05 divided by the number of  $t$ -tests conducted, which should be equal to the number of all dependent variables) [107]. Results indicated that the self-assessment and peer-assessment group were statistically comparable for their social factors (i.e., age, years of English learning and gender), writing performance, and rating consistency in self-assessment of writing prior to the intervention. With no significant difference detected, therefore, further analyses were feasible on the basis of their equivalent conditions.

## 4. Results

### 4.1. Comparisons of the Intervention and the Comparison Groups in the Pre-Test

With assumptions of the *t*-test satisfactorily met, the potential participants' differences concerning social contextual factors (i.e., age and years of English learning), writing performance, and rating accuracy in self-assessment of writing between the two groups at the pre-test were evaluated using a series of independent-samples *t*-tests. To protect against type I errors, the Bonferroni procedure was applied to each *t*-test with the adjusted alpha (0.05 divided by the number of *t*-tests conducted, which should be equal to the number of all dependent variables) [107].

Adjusting the Bonferroni value at 0.0025, we can see that the results showed no significant differences between groups in the two conditions (self-assessment and peer-assessment) regarding age and years of learning English (students' average ages:  $M_{SA} = 19.50$ ,  $SD = 0.67$  and  $M_{PA} = 19.54$ ,  $SD = 0.67$ ; years of English learning on average:  $M_{SA} = 11.55$ ,  $SD = 0.99$  and  $M_{PA} = 11.84$ ,  $SD = 0.77$ ). Students' gender between two conditions was compared and checked by a chi-square test of independence, with no statistically significant difference detected,  $\chi^2(1) = 1.09$ ,  $p = 0.30$ .

Similarly, the conduct of independent-samples *t*-tests revealed no significant difference in the pre-test writing performance between the two groups (self-assessment and peer assessment) before the intervention with adjusted Bonferroni value at 0.001.

Students' initial rating accuracy in self-assessment of writing was measured by strength of the correlation (indicated by Pearson correlation coefficient: *r*) between the students' rating and the raters' rating on writing samples (four individual dimensions and the overall performance) from the pre-test [104]. Following Cohen [105,106], we report the effect size of the Person correlation coefficient *r*, with  $r = 0.1$ ,  $r = 0.3$ , and  $r = 0.5$  representing small, medium, and large effect sizes, respectively.

Overall, in the pre-test, students' ratings of both groups showed a low correlation with the raters' rating for most dimensions (three out of four) (from  $r = 0.11$  to  $r = 0.26$ , and only the dimension of language resources was the exception in which students and the raters' ratings in both groups were negatively related. With regard to the overall performance, students' and the raters' rating in both groups had moderate correlations (intervention group:  $r = 0.34$ ; comparison group:  $r = 0.45$ ).

Statistically, the self-assessment and peer assessment group were comparable for their social factors, writing performance, and their rating accuracy in self-assessment of writing prior to the intervention. Therefore, further analyses were feasible according to such comparable conditions.

### 4.2. Effects of Self-Assessment-Based Intervention on Writing Performance

To check whether the use of self-assessment could improve students' writing performance, students' writing quality was examined from four dimensions. Table 3 shows that students from both the intervention (self-assessment) and the comparison groups (peer-assessment) experienced a rise in writing scores in task achievement, coherence and cohesion, language resources, mechanics, and overall performance. Adjusting the Bonferroni value at 0.001, and according to the Cohen's *d* levels shown, the effect of improvement was strong at the post-test for both groups, with the intervention group outperforming the comparison group in task achievement, coherence and cohesion, mechanics, and overall writing scores in different magnitude. Specifically, students from the intervention group achieved greater significant gains (Cohen's  $d = 1.89$ ) in their overall writing scores than the comparison group (Cohen's  $d = 0.78$ ).

**Table 3.** Descriptive statistics and results of paired-samples *t*-tests of writing scores at the pre- and post-tests in the intervention group (SA) and the comparison group (PA).

Writing Dimensions	Group	Pre-Test (T1)			Post-Test (T2)			T1 vs. T2		
		N	M	SD	N	M	SD	<i>t</i>	<i>p</i>	Cohen's <i>d</i>
Task achievement	SA	51	3.18	0.51	51	4.02	0.45	−11.68	<0.001	1.62
	PA	41	3.04	0.60	41	3.74	0.42	−8.33	<0.001	1.32
Coherence and cohesion	SA	51	3.17	0.51	51	3.96	0.46	−11.08	<0.001	1.54
	PA	41	3.13	0.40	41	3.63	0.44	−9.39	<0.001	1.49
Language resources	SA	51	3.07	0.48	51	3.59	0.50	−7.65	<0.001	1.06
	PA	41	3.12	0.52	41	3.59	0.50	−6.49	<0.001	1.03
Mechanics	SA	51	2.89	0.52	51	3.43	0.51	−7.65	<0.001	1.06
	PA	41	2.82	0.40	41	3.10	0.43	−4.78	<0.001	0.76
Total score	SA	51	12.30	1.55	51	14.99	1.57	−13.63	<0.001	1.89
	PA	41	12.11	1.37	41	13.77	2.26	−4.94	<0.001	0.78

Note. The writing rubric measured four dimensions of writing performance, using a five-scale scoring scheme (1–5) for each dimension.

Furthermore, according to Table 3, the effects of using self-assessment or peer assessment on writing are contingent upon the specific language domain examined. For example, the intervention appears to have had a larger effect on task achievement than coherence and cohesion, as well as mechanics. Moreover, results from independent-samples *t*-tests demonstrated that the writing performance for the intervention group and the comparison group was mostly clearly differentiated ( $p < 0.001$ ) at the post-test. However, for language resources, although both groups registered within-group writing performance improvement (see Table 3) in the post-test, they did not exhibit any between-group differences (see Table 4).

**Table 4.** Descriptive statistics and independent-samples *t*-tests of writing test scores between the intervention group (SA) and the comparison group (PA) at the post-test.

Writing Test Score	Condition	N	M	SD	<i>t</i>	<i>p</i>	95% CI	
							LL	UL
Task achievement	SA	51	4.04	0.43	3.59	0.001	0.15	0.50
	PA	41	3.72	0.43				
Coherence and cohesion	SA	51	3.97	0.46	3.65	0.000	0.16	0.53
	PA	41	3.62	0.44				
Language resources	SA	51	3.58	0.50	−0.24	0.810	−0.23	0.18
	PA	41	3.60	0.50				
Mechanics	SA	51	3.44	0.51	3.41	0.001	0.14	0.54
	PA	41	3.10	0.42				
Total scores	SA	51	15.03	1.57	3.18	0.002	0.47	2.05
	PA	41	13.76	2.23				

Note. CI = confidence interval; LL = lower limit; UL = upper limit.

#### 4.3. Effects of Self-Assessment-Based Intervention on Students' Rating Accuracy in Self-Assessment of Writing

Students' rating accuracy in self-assessment of writing was measured not only quantitatively by the Pearson correlation coefficient (*r*) concerning four individual dimensions (i.e., task achievement, coherence and cohesion, language resources, and mechanics) and the overall performance of EFL writing, as well as qualitatively through evaluating sample

students' and teachers' comments on the selected dimension of students' writing strengths and weaknesses. The strength of students' and raters' correlations in different dimensions in the post-writing test are displayed in Table 5 using bivariate Pearson correlations with two-tailed significance.

**Table 5.** Correlations of students' and the raters' rating on post-writing test samples by the intervention group and the comparison group.

Variables	TA		CC		LR		M		Overall	
	Effect Size <i>r</i>									
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
SR vs. RR (intervention)	0.19	0.38 **	0.26	0.13	−0.06	0.25	0.24	0.11	0.34 **	0.38 **
SR vs. RR (comparison)	0.23	0.12	0.25	0.13	−0.15	0.29	0.11	0.01	0.45 **	0.42 **

Note. TA = task achievement; CC = coherence and cohesion; LR = language resources; M = mechanics; SR = students' rating; RR = raters' rating; \*\*  $p < 0.01$ .

For both groups, according to the bivariate correlations, the students' rating associated with the raters' rating positively for the four writing dimensions and with the overall performance in the post-test. It is notable that, regarding the dimension of language resources, while the students' and the raters' ratings in both groups were negatively associated in the pre-test, the students' and the raters' ratings showed a positive correlation in the post-test, despite the fact that it was a nonsignificant small effect (SA:  $r = 0.25$ ; PA:  $r = 0.29$ ). However, students' and raters' ratings for the dimension of task achievement for the intervention group were significantly positively correlated ( $r = 0.38$ ). In contrast, the correlation between the students' and the raters' ratings for other dimensions decreased in different extent, for both groups in the post-test. For the overall performance, similar to the pre-writing test, students' and the raters' ratings in both groups had significant moderate positive correlations at the post-writing test. Although remaining at a medium level, the students' and the raters' rating correlations for the intervention group indicated a small growth (from  $r = 0.34$  to  $0.38$ ); in contrast, the comparison group had a slight decrease (from  $r = 0.45$  to  $0.42$ ).

When commenting on students' writing strengths and weaknesses, raters and students were asked to indicate, first, a writing dimension (e.g., task achievement) and, then, for each student's writing, why it was a strong or weak point. Although, as seen from the above, there were low correlations between the students' rating and the raters' rating, increasing similarity between the students' and the raters' comments from pre- to post-writing tasks suggests that self-assessment of writing had a positive effect on students' rating accuracy (for details of six students' and the raters' comments in the pre- and post-writing tasks, see Appendix C).

Specifically, in the pre-test, considerable discrepancies were found between students' and the raters' comments, suggesting that teachers and students hold diverse ideas of a student's strengths and weaknesses in writing. Nonetheless, in the post-test, the students' and the raters' comments overlapped in both the designated rubric dimensions and the comments on those dimensions. Students' comments were also more expressive and reflective in the post-test than in the pre-test.

In summary, correlations of the students' and the raters' ratings for both groups were significant only in overall writing performance and task achievement for the intervention group at the post-test. No significant associations were found in coherence and cohesion, language resources, and mechanics. The results imply students' low rating accuracy in terms of self-assessment of individual writing dimensions, but relatively higher, intermediate, rating accuracy in self-assessing their overall writing performance.



## 5. Discussion

Informed by formative assessment and self-regulated learning theories, the current study was conducted to investigate the effects of using self-assessment on EFL students' writing performance and rating accuracy. Consistent with previous studies in which the positive effects of self-assessment on overall writing quality improvement have been demonstrated [43,72], this study revealed that students' writing performance improved in both intervention and comparison groups, and the increase in the intervention group's (self-assessment) writing performance had a greater effect size (Cohen's  $d = 1.89$ ) than that of the comparison group (peer-assessment, Cohen's  $d = 0.78$ ) in the writing dimensions of task achievement, coherence and cohesion, and mechanics. The qualitative data complemented the quantitative findings, with all six students indicating positive formative development in their rating accuracy.

It is worth noting that both groups displayed within-group significant growth (Cohen's  $d = 1.06$  and  $1.03$  for the intervention group and comparison group, respectively) but no between-group differences in the post-test in language resources, the only dimension in which this was evident. One conceivable explanation of this might be that, during the intervention, the teacher's instruction did not emphasise students' use of vocabulary and sentence structures. Hence, the results suggest that self-assessment and peer-assessment influenced students' use of language resources to a similar extent. Potentially, it is possible that the task selection and design of this study might have been cognitively demanding for students to make use of sophisticated language to respond to the task within a short amount of time, especially when self-assessment was not familiar to them [108,109].

In terms of students' rating accuracy, only moderately significant correlations were found between the students' and the raters' ratings for both groups' overall writing performance, with very little change from the pre-test to the post-test in the correlations. The nonsignificant correlations for coherence and cohesion, language resources, and mechanics reduced further in the post-test. Such results also agree with earlier studies [23,64,67], indicating that, when self-assessment practices are learning-oriented, students are able to judge their overall writing similarly to external raters, but rating discrepancies may exist for the individual dimensions. Even though there was a significant correlation ( $r = 0.38$ ) between the students' and the raters' rating in the dimension of task achievement, in the other individual writing dimensions, the correlation between the students' and the raters' rating decreased in the post-test. This result partially corroborates Liu's finding [95], in which Chinese students' ratings correlated with the raters' only in task completion due to their limited English proficiency to rate other writing dimensions accurately, and that students need extra time and practice to make their rating comparable to those made by the raters.

There are a few plausible explanations for the reduced correlation between the students' rating and the raters' rating for individual writing dimensions (except for task achievement). Initially, it may be the effect of engaging in self-assessment, in which students report negatively on themselves when reflecting on their self-assessment over a period of time. This can lead students to self-doubt and unnecessary anxiety [4,71,110], negatively affecting their judgement of their work. Secondly, as indicated in the previous literature, students with different language proficiency levels [30,95] may either underestimate or overestimate themselves in self-assessment because of ego protection or lack of relevant knowledge [12,28,36,74,76]. Thirdly, learners in this study, as novice raters, were probably given insufficient classroom instructions and time to digest the self-assessment or peer-assessment materials and practices. Therefore, they were likely less ready to assess their work, such as vocabulary and sentence structures in writing, using the detailed rubric provided, than the experienced raters [15,64]. Another reason may be that the students had inadequate teacher feedback during the intervention for more accurate student self-assessment development. At this early stage, learners may have still lacked proper understanding and sufficient experience in self-assessment [8].

Nevertheless, from a formative perspective, students' rating accuracy might have enhanced after self-assessment as students' comments on writing strengths and weaknesses are more in line with the raters' [8]. Those findings broadly support the statement that "self-assessment may not lead to an improved essay; however, it may lead to an enhanced insight into the strengths and weaknesses of the essay the students will submit" [111] (p. 65). According to Hattie and Timperley's four levels of feedback focus [112], the fact that the six students were able to generate more precise, observant, and constructive comments on their writing strengths and weaknesses in the post-test supports the argument that self-assessment can support students to refine their judgements, raise their awareness of their writing quality, and nurture and sustain their writing development [25,113–120]. Therefore, the question is not about how faithful students can be in self-assessing their writing, but about how the engagement in self-assessing their performance can foster sustainable writing development. Arguably, using self-assessment could be one approach to realise that goal [63].

## 6. Research Implications

This study affords some implications for EFL writing researchers to further investigate issues related to self-assessment of writing for better implementation of self-assessment in their own contexts, especially in the Chinese context or contexts that share similarities.

Theoretically, evidence gained from the positive effects of implementing self-assessment supports the effectiveness of rubric-referenced self-assessment and reflects students' abilities to construct and self-regulate knowledge with teachers' constant feedback and scaffolding [36,121,122]. Methodologically, the quasi-experimental design of this study has the potential to inform, and to be used for, valid and in-depth investigations in similar settings as most previous research intends to adopt either the quantitative or the qualitative approach to investigate the role of self-assessment in the writing domain. Pedagogically, the empirical evidence reported here provides information for the integration of self-assessment as a regular element into curriculum/course design in EFL writing contexts such as China [10,35], where students are accustomed to a centralised education system and tend to accept so-called "standard" or authoritative evaluations [15,25], and doing so will help them develop their self-regulated learning capacity in trying self-assessment.

## 7. Conclusions

This study presents only the tip of a very promising iceberg, i.e., students' self-assessment in the EFL writing context. With the overarching aim of exploring the effect of using self-assessment on EFL students' writing performance and rating accuracy, this study was conducted with 94 participants (92 students and two English lecturers). The empirical evidence provided in previous sections shows that self-assessment can be oriented to foster and enhance students' writing performance by assisting students' active engagement in learning. Nevertheless, given that self-assessment was a novel experience for both the lecturer and the Chinese students in this study, the inclusion of self-assessment in the EFL writing pedagogy may take longer to have a further positive impact on certain aspects, e.g., students' rating accuracy improvement, given the misalignment between students' and raters' rating [85,114]. Although students in this study seemed unable to assess their work reliably in relation to the raters' scores after the intervention, it is still necessary to include self-rating during self-assessment because learning to self-assess not only helps students to reflect on their work, but also sustains their assessing abilities and writing development.

The generalisability of the current research findings is subject to several limitations. The first set of limitations arises from the self-assessment intervention, in which the research site made an unforeseen decision to use peer assessment instead of using the previously agreed traditional teacher assessment in the comparison group. Hence, the current study may not be able to make a strong claim on the effects of using self-assessment in contrast to traditional assessment practices. Comparing the effects of using self-assessment and teacher assessment on students' writing performance in EFL writing settings, particularly

to ensure the implementation of high-quality self-assessment in writing programmes [115], can be a productive area for future research. Many researchers have highlighted that sophisticated self-assessment takes time, and the intervention time for this study may have been inadequate to document more systematic changes in the self-assessment process. Longitudinal studies over a much longer time frame could be helpful to provide richer insights into the effectiveness of the self-assessment intervention [53,116–119].

Another limitation of the current study concerns its sample size. Firstly, as the lecturer participants and student participants were only recruited from one Chinese university through convenience sampling, cautions need to be taken when interpreting the research findings. It is also acknowledged that the qualitative data that six students provided are not intended for making generalisations of students' formative rating accuracy development in self-assessment of writing across other populations, as the representation is limited. What merits future research attention is to expand the sampling strategies and sample size so as to engage more students of different majors from the same university or students of similar majors from different universities. In that sense, the varied representation of population in self-assessment of writing research can possibly be ensured. More in-depth qualitative studies adopting different methodologies to enrich the perception of students' lived experiences in self-assessment of writing, are anticipated. For instance, more attention could also be given to interpreting the mental and emotional reactions that students experience during self-assessment so that teachers could better support students to develop effective forms of self-assessment to maximise the benefits gained from the self-assessment of writing practices.

Commenting on Black and Wiliam's work [22], Lee argued that "self-assessment by the student is not an interesting option or luxury; it has to be seen as essential" [19] (p. 55), which emphasised that the idea of self-assessment ought to be an essential practice integrated into every writing classroom to promote sustainable learning [120–123], instead of being treated as an occasional practice. It is expected that findings of this study will encourage practitioners to implement and monitor self-assessment in their classroom despite the challenges and difficulties that may arise in the pedagogical process.

**Author Contributions:** X.S.Z. and L.J.Z. conceptualised and designed the study; X.S.Z. collected and analysed the data and wrote the first draft; all authors revised the manuscript before L.J.Z. finalised and submitted it as the corresponding author. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Human Participants Ethics Committee of the University of Auckland (Protocol No. 019574, approved 2 August 2017).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

For the pre-test (45 min)

Prompts: Shopping online is a new trend throughout the world today, and the reasons why people choose to shop online is complex. Such a trend may transform people's lives, or even how the society operates.

Write an essay with at least 200 words on the following topic: What are the reasons for people shopping online and what effects does it have on people's lives and the society?

---

---

---

---

---

---

---

---

---

---

For the post-test (45 min)

Prompts: Even though shopping online is popular around the world, a large number of people still prefer shopping in physical stores due to different reasons, and how shopping in physical stores could shape people's lives and the whole society is not clear.

Write an essay with at least 200 words on the following topic: What are the reasons for people shopping in physical stores and what effects does it have on people's lives and the society?

---

---

---

---

---

---

---

---

---

---

## Appendix B

There are four sections in this writing rubric, namely, task achievement, coherence and cohesion, language resources, and mechanics. Within each section are descriptions of characteristics representing writing at different levels, from 1–5, where 5 represents the highest level of quality. I would like to request you to self-assess your writing according to the rubric, and generate a score for yourself in each section and an average overall score. Lastly, please write down your comments on the strengths and weaknesses of your writing. You have 20 min to complete this task. The meanings or definitions of the key dimensions in the writing rubric are provided below. I suggest that you read through the rubric to become familiar with it before starting. Thank you very much for your generous help.

- Task achievement: This is the extent to which you responded to all parts of the topic questions, and how you developed relevant ideas to support them with explanations, examples, or experiences. It is important that your opinion is clear and relevant to the topic.
- Coherence and cohesion: Coherence means that you presented your ideas in a logical way, and cohesion refers to the degree you connected sentences (or different parts of one sentence) and paragraphs with linking words (e.g., since, in addition, because, first, although, however, and moreover) such that the reader could follow your ideas easily. Both elements build up the organisation of your writing.
- Language resources: This refers to your ability to use a variety of vocabulary (e.g., different word forms: noun/verb, synonyms, phrases, formal expressions), grammar (e.g., different tenses: I had gone/I will be going, and relative clauses: who/which/that etc.), and sentence structures (e.g., simple sentences—one main clause/one verb; complex sentences—at least one independent clause plus at least one dependent clause linked by because, although, etc.; compound sentences—two or more independent clauses joined by and, but, or, etc.) in a balanced, flexible and accurate way in your writing.

- **Mechanics:** This is your ability to apply correct spelling, punctuation (e.g., comma, period, and question mark), capitalisation, abbreviations (e.g., DIY and CEO), paragraphing (divide your writing into logical parts), and grammar rules (e.g., the use of the/a/an, uncountable nouns, plural forms, use of phrases match, and avoiding run-on sentences, in which two or more independent clauses (i.e., complete sentences) are joined without an appropriate conjunction or mark of punctuation) in your writing.

I believe my strengths in this writing are:

I think my weaknesses in this writing are:

My scores are • task achievement, • coherence and cohesion, • language resources, and • mechanics. My general score is . . .

**Table A1.** Self-assessment of writing rubric.

Level	Task Achievement	Coherence and Cohesion	Language Resources	Mechanics
5	<ul style="list-style-type: none"> <li>• Proficiently paraphrase and address all the task questions.</li> <li>• Clearly express your opinions through the task and support them with at least four relevant, detailed examples, and personal experiences. Meet or moderately exceed the word limit.</li> </ul>	<ul style="list-style-type: none"> <li>• Ideas, sentences, and paragraphs are clearly and logically presented and linked by a variety of linking words accurately.</li> <li>• With topic and closing sentences, the whole writing has a nice flow and is pleasant to read.</li> </ul>	<ul style="list-style-type: none"> <li>• Use a wide range of topic-related and sophisticated vocabulary naturally to answer task questions; may have rare mistakes in word choice, and forms.</li> <li>• Use a wide variety of sentence structures, and all are used in a flexible and correct way.</li> </ul>	<ul style="list-style-type: none"> <li>• No errors in spelling, punctuation, capitalisation, grammar rules.</li> <li>• Divide the writing into reasonable paragraphs, and each paragraph expresses a clear and logical idea and indicates a change of focus at the beginning. Clear and neat handwriting.</li> </ul>
4	<ul style="list-style-type: none"> <li>• Appropriately paraphrase and address about 85% of the task questions.</li> <li>• Clearly express your opinions through the task and support them with three generally relevant, detailed examples, and personal experiences, but they may lack focus sometime. Meet or slightly exceed the word limit.</li> </ul>	<ul style="list-style-type: none"> <li>• Though 1–2 expressions are unclear, ideas, sentences, and paragraphs are clearly and logically presented in general. Some linking words are underused or overused.</li> <li>• With topic and closing sentences, the whole writing is easy to follow.</li> </ul>	<ul style="list-style-type: none"> <li>• Use a wide range of topic-related and sophisticated vocabulary to answer task questions; may have 1–2 mistakes in word choice, and word forms.</li> <li>• Use a wide variety of sentence structures, and about 85% of them are error-free sentences.</li> </ul>	<ul style="list-style-type: none"> <li>• 1–3 errors in spelling, punctuation, capitalisation, grammar rules.</li> <li>• Divide the writing into reasonable paragraphs, and each paragraph expresses a central idea and indicates a change of focus at the beginning. Clear and neat handwriting.</li> </ul>



Table A1. Cont.

Level	Task Achievement	Coherence and Cohesion	Language Resources	Mechanics
3	<ul style="list-style-type: none"> <li>Clearly address and paraphrase around half of the task questions.</li> <li>Express a relevant but unclear personal opinion through the task and support it with at least one relevant, detailed example and personal experience. Slightly or moderately below the word limit.</li> </ul>	<ul style="list-style-type: none"> <li>Parts of ideas, sentences and paragraphs are repetitive and illogical. Different linking words are used, but there are 3–4 mistakes within/between sentences that may block reading but not affect communication.</li> <li>The writing has topic sentences for each part, but no conclusion.</li> </ul>	<ul style="list-style-type: none"> <li>Use basic topic-related vocabulary to answer task questions, but 3–5 errors in word forms or grammar of uncommon words.</li> <li>Use a mix of simple and complex sentence structures; simple sentence structures are used correctly, while having 2–4 mistakes in complex sentences.</li> </ul>	<ul style="list-style-type: none"> <li>4–7 noticeable errors in spelling, punctuation, capitalisation, grammar rules.</li> <li>Divide the writing into reasonable paragraphs, but each paragraph expresses more than one idea. Good handwriting without affecting reading.</li> </ul>
2	<ul style="list-style-type: none"> <li>Unable to address any part of the task questions, and almost copy all the words of the task questions.</li> <li>Express an unclear personal opinion through the task but the supporting examples, and personal experiences are irrelevant and not well-explained. Considerably below the word limit.</li> </ul>	<ul style="list-style-type: none"> <li>It is hard to follow the logic in ideas, sentences, and paragraphs. Only basic linking words (e.g., and/or/but/first/however) are used, and they may be used incorrectly and repetitively.</li> <li>The writing does not have a clear topic sentence and a conclusion.</li> </ul>	<ul style="list-style-type: none"> <li>Use only basic vocabulary and some of them are not related to the task. Over 6 errors in word forms and the meaning of ideas may be changed by the errors.</li> <li>Use only simple sentences without clauses. Many incomplete and run-on sentences.</li> </ul>	<ul style="list-style-type: none"> <li>Over 8 errors in spelling, punctuation, capitalisation, grammar rules that the writing is hard to read and understand.</li> <li>There are fewer than 3 paragraphs or more than 5 paragraphs. Ideas for different questions are mixed, so the meaning is confusing for the reader. Poor handwriting, difficult to read.</li> </ul>
1	<ul style="list-style-type: none"> <li>The answer is mostly unrelated to the task, no personal opinion is expressed, and it seems part of the writing is a memorised response. Extremely below the word limit.</li> </ul>	<ul style="list-style-type: none"> <li>No logic in the presented ideas, sentences, and paragraphs; wrong use of all the linking words; no beginning or ending.</li> </ul>	<ul style="list-style-type: none"> <li>Use extremely limited and repetitive vocabulary, and most of the word forms are used wrongly; only run-on sentences or phrases which are hard to read.</li> </ul>	<ul style="list-style-type: none"> <li>With very poor handwriting, the writing is full of errors and no paragraph formatting. Almost impossible to read.</li> </ul>

## Appendix C

Table A2. Sample of students and teachers' comments of writing.

CommentSources	Strengths		Weaknesses	
	Pre-Comments	Post-Comments	Pre-Comments	Post-Comments
Student A	Language resources: I used complicated sentence structures.	Task achievement: I answered the topic questions clearly with good examples in each part. I also have a clear structure with signal words linking my ideas.	Mechanics: Poor spelling.	Language resources: I am not certain if I used enough authentic expressions and complex structures in the right way in my writing. I need to work on various sentence structures.

Table A2. Cont.

CommentSources	Strengths		Weaknesses	
	Pre-Comments	Post-Comments	Pre-Comments	Post-Comments
Teachers	Task achievement: Answered most task questions with clear opinions and examples.	Task achievement and language resources: Addressed all task questions with relevant examples and good logic. Also used sophisticated expressions and attributive clauses in the right way.	Coherence and cohesion: Lack of linking words to link ideas smoothly.	Language resources: Run-on sentences in complex sentence structures.
Student B	Task achievement: I have abundant ideas for the topic.	Task achievement and language resources: I answered all the task questions with proper examples to support my arguments. I tried to use a range of academic expressions.	Language resources: Not good at using complex vocabulary /	Mechanics: I made spelling mistakes because I focused more on the words and sentence structures I used, and I neglected word spelling. I need to make time to double-check spelling alone.
Teachers	Coherence and cohesion/ language resources: The writing is a pleasure to read with ideas clearly and logically presented. Also used a variety of sentence structures and sophisticated vocabulary.	Task achievement and language resources: Paraphrased and addressed all the task questions with detailed examples and used sophisticated vocabulary and structure.	Mechanics: Violated many grammar rules.	Mechanics: Numerous spelling errors.
Student C	Coherence and cohesion: Logical thoughts.	Coherence and cohesion: My writing has topic sentences for each part, but no conclusion. The whole writing is easy to follow with enough examples.	Mechanics: Punctuation mistakes.	Mechanics: I realise I always used tenses wrongly. I guess I do not fully understand the occasions to use the present tense. My handwriting is not easy to follow, and I should start from improving my handwriting.
Teachers	Language resources: Generally correct use of complex sentence structures.	Coherence and cohesion: Ideas are presented clearly and logically with good use of linking words.	Task achievement: Unable to answer the task questions; extremely below the word limit.	Mechanics: Many mistakes in grammar, especially the use of tense. Poor handwriting disturbs reading.
Student D	Mechanics: Very few spelling mistakes.	Language resources: To answer the task questions, I tried to use different kinds of words in different forms to achieve vocabulary variety.	Mechanics: I always spelled words wrong.	Coherence and cohesion: My ideas in the body part were not logical and coherent for the topic. I need to make stronger links among my ideas and examples using transitional words.
Teachers	Task achievement: Clearly answered task questions with enough examples.	Language resources: Used a range of words in the right way to address task questions.	Language resources: Major problems in using the right word/phrase collocation.	Coherence and cohesion: Repetitive and illogical ideas; not easy to follow.

Table A2. Cont.

CommentSources	Strengths		Weaknesses	
	Pre-Comments	Post-Comments	Pre-Comments	Post-Comments
Student E	I do not know.	Task achievement: I have a clear opinion, and I can write effective sentences to describe my examples and to support my ideas.	Task achievement: Not sure about my choice.	Mechanics: I always forget spelling and grammar rules during writing. I need more practice in grammar such as subject and verb agreement.
Teachers	Language resources: Used a range of topic-related words.	Task achievement: Most task questions are addressed with supporting details.	Mechanics: Violated many basic grammar rules.	Mechanics: Full of spelling errors and grammar mistakes.
Student F	Task achievement: I do not know why.	Task achievement: I used four examples to express myself clearly around the main topic. I think I provided enough support for my argument.	No idea.	Language resources: I used simple sentence structures and basic vocabulary too often, such as “think, like, etc.”. I need to learn to use more advanced words.
Teachers	Coherence and cohesion: Opinions are clear in general.	Task achievement: Most task questions are answered with relevant examples.	Task achievement: Half of the task questions are not addressed.	Language resources: Mainly used basic words and sentence structures with phrase collocation errors.

## References

- Cassidy, S. Assessing “inexperienced” Students’ Ability to Self-Assess: Exploring Links with Learning Style and Academic Personal Control. *Assess. Eval. High. Educ.* **2007**, *32*, 313–330. [\[CrossRef\]](#)
- Dochy, F.; Segers, M.; Sluijsmans, D. The Use of Self-, Peer and Co-Assessment in Higher Education: A Review. *Stud. Higher Education* **1999**, *24*, 331–350. [\[CrossRef\]](#)
- Sambell, K.; McDowell, L.; Montgomery, C. Developing Students as Self-Assessors and Effective Lifelong Lea. In *Assessment for Learning in Higher Education*; Routledge: New York, NY, USA, 2013; pp. 120–146.
- Yan, Z. Self-Assessment in the Process of Self-Regulated Learning and Its Relationship with Academic Achievement. *Assess. Eval. High. Educ.* **2020**, *45*, 224–238. [\[CrossRef\]](#)
- Boud, D. *Enhancing Learning through Self-Assessment*; Routledge/Falmer: New York, NY, USA, 1995.
- Boud, D.; Soler, R. Sustainable Assessment Revisited. *Assess. Eval. High. Educ.* **2016**, *41*, 400–413. [\[CrossRef\]](#)
- Earl, L.; Katz, S. *Rethinking Classroom Assessment with Purpose in Mind: Assessment for Learning, Assessment as Learning, Assessment of Learning*; Manitoba Education, Citizenship and Youth: Winnipeg, MB, Canada, 2006.
- González-Betancor, S.M.; Bolívar-Cruz, A.; Verano-Tacoronte, D. Self-Assessment Accuracy in Higher Education: The Influence of Gender and Performance of University Students. *Act. Learn. High. Educ.* **2019**, *20*, 101–114. [\[CrossRef\]](#)
- Bourke, R. Self-Assessment in Professional Programmes within Tertiary Institutions. *Teach. High. Educ.* **2014**, *19*, 908–918. [\[CrossRef\]](#)
- Nielsen, K. Peer and Self-Assessment Practices for Writing across the Curriculum: Learner-Differentiated Effects on Writing Achievement. *Educ. Rev.* **2019**, *73*, 753–774. [\[CrossRef\]](#)
- Ross, J.A.; Rolheiser, C.; Hogaboam-Gray, A. Effects of Self-Evaluation Training on Narrative Writing. *Assess. Writ.* **1999**, *6*, 107–132. [\[CrossRef\]](#)
- Sullivan, K.; Hall, C. Introducing Students to Self-Assessment. *Assess. Eval. High. Educ.* **1997**, *22*, 289–305. [\[CrossRef\]](#)
- Andrade, H.L. A Critical Review of Research on Student Self-Assessment. *Front. Educ.* **2019**, *4*, 87. [\[CrossRef\]](#)
- Nicol, D.J.; Macfarlane-dick, D. Formative Assessment and Self-Regulated Learning: A Model and Seven Principles of Good Feedback Practice. *Stud. High. Educ.* **2006**, *31*, 199–218. [\[CrossRef\]](#)
- Xu, Y. Scaffolding Students’ Self-Assessment of Their English Essays with Annotated Samples: A Mixed-Methods Study. *Chin. J. Appl. Linguist.* **2019**, *42*, 503–526. [\[CrossRef\]](#)
- Andrade, H.L.; Brown, G.T. Student Self-Assessment in the Classroom. In *Handbook of Human and Social Conditions in Assessment*; Brown, G.T.L., Harris, L.R., Eds.; Routledge: New York, NY, USA, 2016; pp. 319–334.
- Panadero, E.; Broadbent, J.; Boud, D.; Lodge, J.M. Using Formative Assessment to Influence Self- and Co-Regulated Learning: The Role of Evaluative Judgement. *Eur. J. Psychol. Educ.* **2019**, *34*, 535–557. [\[CrossRef\]](#)

18. Tai, J.; Ajjawi, R.; Boud, D.; Dawson, P.; Panadero, E. Developing Evaluative Judgement: Enabling Students to Make Decisions about the Quality of Work. *High. Educ.* **2018**, *76*, 467–481. [\[CrossRef\]](#)
19. Lee, I. *Classroom Writing Assessment and Feedback in L2 School Contexts*; Springer: Berlin/Heidelberg, Germany, 2017.
20. Nielsen, K. Self-Assessment Methods in Writing Instruction: A Conceptual Framework, Successful Practices and Essential Strategies. *J. Res. Read.* **2012**, *37*, 1–16. [\[CrossRef\]](#)
21. Sadler, R. Formative Assessment and the Design of Instructional Systems. *Instr. Sci.* **1989**, *18*, 119–144. [\[CrossRef\]](#)
22. Black, P.; Wiliam, D. Assessment and Classroom Learning. *Assess. Educ. Princ. Policy Pract.* **1998**, *5*, 7–74. [\[CrossRef\]](#)
23. Longhurst, N.; Norton, L.S. Self-Assessment in Coursework Essays. *Stud. Educ. Eval.* **1997**, *23*, 319–330. [\[CrossRef\]](#)
24. Tan, K.H.K. Does Student Self-assessment Empower or Discipline Students? *Assess. Eval. High. Educ.* **2004**, *29*, 651–662. [\[CrossRef\]](#)
25. Chen, Y.-M. Learning to Self-Assess Oral Performance in English: A Longitudinal Case Study. *Lang. Teach. Res.* **2008**, *12*, 235–262. [\[CrossRef\]](#)
26. Crooks, T. Assessment for Learning in the Accountability Era: New Zealand. *Stud. Educ. Eval.* **2011**, *37*, 71–77. [\[CrossRef\]](#)
27. Dann, R. Assessment as Learning: Blurring the Boundaries of Assessment and Learning for Theory, Policy and Practice. *Assess. Educ. Princ. Policy Pract.* **2014**, *21*, 149–166. [\[CrossRef\]](#)
28. Pintrich, P.R. The Role of Metacognitive Knowledge in Learning, Teaching, and Assessing. *Am. J. Psychol.* **2002**, *41*, 219–225. [\[CrossRef\]](#)
29. Zhang, X.S.; Zhang, L.J.; Parr, J.M.; Biebricher, C. Exploring Teachers' Attitudes and Self-Efficacy Beliefs for Implementing Student Self-Assessment of English as a Foreign Language Writing. *Front. Psychol.* **2022**, *13*, 1–15. [\[CrossRef\]](#)
30. Harris, L.R.; Brown, G.T.L. Why Use Self-Assessment in the Classroom? In *Using Self-Assessment to Improve Student Learning*; Routledge: New York, NY, USA, 2018; pp. 1–14.
31. Huang, S.C. Understanding Learners' Self-Assessment and Self-Feedback on Their Foreign Language Speaking Performance. *Assess. Eval. High. Educ.* **2016**, *41*, 803–820. [\[CrossRef\]](#)
32. Panadero, E.; Jonsson, A.; Strijbos, J.-W. Scaffolding Self-Regulated Learning through Self-Assessment and Peer Assessment: Guidelines for Classroom Implementation. In *Assessment for Learning: Meeting the Challenge of Implementation*; Laveault, D., Allal, L., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 311–324. [\[CrossRef\]](#)
33. Sadek, N. The Effect of Self-Assessment as a Revision Technique on Egyptian EFL Students' Expository Essay Writing. In *Assessing EFL Writing in the 21st Century Arab World: Revealing the Unknown*; Ahmed, A., Abouabdelkader, H., Eds.; Palgrave Macmillan: Cham, Switzerland, 2018; pp. 21–52. [\[CrossRef\]](#)
34. Oscarson, M. Self-Assessment of Language Proficiency: Rationale and Applications. *Lang. Test.* **1989**, *6*, 1–13. [\[CrossRef\]](#)
35. Tan, K. Conceptions of Self-Assessment: What Is Needed for Long-Term Learning? In *Rethinking Assessment in Higher Education: Learning for the Long Term*; Boud, D., Falchikov, N., Eds.; Routledge: London, UK, 2007; pp. 114–127.
36. Zhang, L.J. A Dynamic Metacognitive Systems Account of Chinese University Students' Knowledge about EFL Reading. *TESOL Q.* **2010**, *44*, 320–353. [\[CrossRef\]](#)
37. Zhang, L.J. L2 Writing: Toward a Theory-Practice Praxis. In *Handbook of Practical Second Language Teaching and Learning*; Hinkel, E., Ed.; Routledge: New York, NY, USA, 2022; pp. 331–343. [\[CrossRef\]](#)
38. Panadero, E.; Brown, G.T.L.; Strijbos, J.W. The Future of Student Self-Assessment: A Review of Known, Unknowns and Potential Directions. *Educ. Psychol. Rev.* **2016**, *28*, 803–830. [\[CrossRef\]](#)
39. Van Reybroeck, M.; Penneman, J.; Vidick, C.; Galand, B. Progressive Treatment and Self-Assessment: Effects on Students' Automatisations of Grammatical Spelling and Self-Efficacy Beliefs. *Read. Writ.* **2017**, *30*, 1965–1985. [\[CrossRef\]](#)
40. Andrade, H.L. Students as the Definitive Source of Formative Assessment: Academic Self-Assessment and the Self-Regulation of Learning. In *Handbook of Formative Assessment*; Andrade, H., Cizek, G.J., Eds.; Routledge: New York, NY, USA, 2010; pp. 90–105.
41. Black, P. Formative Assessment—An Optimistic but Incomplete Vision. *Assess. Educ. Princ. Policy Pract.* **2015**, *22*, 161–177. [\[CrossRef\]](#)
42. Panadero, E.; Alonso-Tapia, J. Self-Assessment: Theoretical and Practical Connotations. When It Happens, How Is It Acquired and What to Do to Develop It in Our Students. *Electron. J. Res. Educ. Psychol.* **2013**, *11*, 551–576. [\[CrossRef\]](#)
43. Andrade, H.L.; Du, Y.; Mycek, K. Rubric-Referenced Self-Assessment and Middle School Students' Writing. *Assess. Educ. Princ. Policy Pract.* **2010**, *17*, 199–214. [\[CrossRef\]](#)
44. Mat, Y.N.; Par, L. Employing a Self-Assessment Rubric on the EFL Students' Writing Activities: Is It Effective? *Engl. Lang. Educ. J.* **2022**, *1*, 1–10.
45. Yan, Z. Student Self-Assessment Practices: The Role of Gender, School Level and Goal Orientation. *Assess. Educ. Princ. Policy Pract.* **2018**, *25*, 183–199. [\[CrossRef\]](#)
46. Andrade, H.; Du, Y. Student Responses to Criteria-referenced Self-assessment. *Assess. Eval. High. Educ.* **2007**, *32*, 159–181. [\[CrossRef\]](#)
47. Andrade, H.; Valtcheva, A. Promoting Learning and Achievement through Self-Assessment. *Theory Into Pract.* **2009**, *48*, 12–19. [\[CrossRef\]](#)
48. Pinner, R. Trouble in Paradise: Self-Assessment and the Tao. *Lang. Teach. Res.* **2016**, *20*, 181–195. [\[CrossRef\]](#)
49. Boud, D. Assessment and Learning: Contradictory or Complementary. In *Assessment and Learning in Higher Education*; Night, P., Ed.; Kogan Page: London, UK, 1995; pp. 35–48.
50. Butler, Y.G.; Lee, J. The Effects of Self-Assessment among Young Learners of English. *Lang. Test.* **2010**, *27*, 5–31. [\[CrossRef\]](#)

51. Brown, G.; Harris, L. The Future of Self-Assessment in Classroom Practice: Reframing Self-Assessment as a Core Competency. *Frontline Learn. Res.* **2014**, *3*, 22–30. [\[CrossRef\]](#)
52. Tan, K.H.K. Meanings and Practices of Power in Academics' Conceptions of Student Self-Assessment. *Teach. High. Educ.* **2009**, *14*, 361–373. [\[CrossRef\]](#)
53. Lee, I. Formative Assessment in EFL Writing: An Exploratory Case Study. *Chang. Engl. Stud. Cult. Educ.* **2011**, *18*, 99–111. [\[CrossRef\]](#)
54. Schunk, D.H.; Usher, E.L. Assessing Self-Efficacy for Self-Regulated Learning. In *Handbook of Self-Regulation of Learning and Performance*; Zimmerman, B.J., Schunk, D.H., Eds.; Routledge: London, UK, 2011; pp. 282–297. [\[CrossRef\]](#)
55. LeBlanc, R.; Painchaud, G.G. Self-Assessment as a Second Language Placement Instrument. *TESOL Q.* **1985**, *19*, 673–687. [\[CrossRef\]](#)
56. Earl, L.M. *Assessment as Learning: Using Classroom Assessment to Maximise Student Learning*, 2nd ed.; Corwin Press: Thousand Oaks, CA, USA, 2013.
57. Ross, J.A. The Reliability, Validity, and Utility of Self-Assessment. *Pract. Assess. Res. Eval.* **2006**, *11*, 10.
58. Rolheiser, C. *Self-Evaluation: Helping Students Get Better at It! A Teacher's Resource Book*; University of Toronto Press: Toronto, ON, Canada, 1996.
59. Burner, T. Formative Assessment of Writing in English as a Foreign Language. *Scand. J. Educ. Res.* **2016**, *60*, 626–648. [\[CrossRef\]](#)
60. Sun, T.; Wang, C. College Students' Writing Self-Efficacy and Writing Self-Regulated Learning Strategies in Learning English as a Foreign Language. *System* **2020**, *90*, 102221. [\[CrossRef\]](#)
61. Xiao, G.; Chen, X. Application of COCA in EFL Writing Instruction at the Tertiary Level in China. *Int. J. Emerg. Technol. Learn.* **2018**, *13*, 160–173. [\[CrossRef\]](#)
62. Zhang, L.J.; Zhang, D. Metacognition in TESOL: Theory and Practice. In *The TESOL Encyclopedia of English Language Teaching*; Lontas, J., Shehadeh, A., Eds.; Wiley: Malden, MA, USA, 2018; pp. 682–792. [\[CrossRef\]](#)
63. Liu, J.; Xu, Y. Assessment for Learning in English Language Classrooms in China: Contexts, Problems, and Solutions. In *Innovation in Language Learning and Teaching: The case of China*; Reinders, H., Nunan, D., Zou, B., Eds.; Palgrave Macmillan: London, UK, 2017; pp. 17–37. [\[CrossRef\]](#)
64. Zhang, J. Same Text Different Processing? Exploring How Raters' Cognitive and Meta-Cognitive Strategies Influence Rating Accuracy in Essay Scoring. *Assess. Writ.* **2016**, *27*, 37–53. [\[CrossRef\]](#)
65. Boud, D.; Falchikov, N. Quantitative Studies of Student Self-Assessment in Higher Education: A Critical Analysis of Findings. *High. Educ.* **1989**, *18*, 529–549. [\[CrossRef\]](#)
66. Kun, A.I. A Comparison of Self versus Tutor Assessment among Hungarian Undergraduate Business Students. *Assess. Eval. High. Educ.* **2016**, *41*, 350–367. [\[CrossRef\]](#)
67. Leach, L. Optional Self-Assessment: Some Tensions and Dilemmas. *Assess. Eval. High. Educ.* **2012**, *37*, 137–147. [\[CrossRef\]](#)
68. Weigle, S.C. *Assessing Writing*; Cambridge University Press: Cambridge, UK, 2002. [\[CrossRef\]](#)
69. Zheng, Y.; Yu, S. What Has Been Assessed in Writing and How? Empirical Evidence from Assessing Writing (2000–2018). *Assess. Writ.* **2019**, *42*, 100421. [\[CrossRef\]](#)
70. Brown, G.T.L.; Harris, L.R. Student Self-Assessment. In *SAGE Handbook of Research on Classroom Assessment*; McMillan, J.H., Ed.; SAGE Publications: Thousand Oaks, CA, USA, 2013; pp. 367–393. [\[CrossRef\]](#)
71. Nawas, A. Grading Anxiety with Self and Peer-Assessment: A Mixed-Method Study in an Indonesian EFL Context. *Issues Educ. Res.* **2020**, *30*, 224–244.
72. Mazloomi, S.; Khabiri, M. The Impact of Self-Assessment on Language Learners' Writing Skill. *Innov. Educ. Teach. Int.* **2018**, *55*, 91–100. [\[CrossRef\]](#)
73. Stefani, L.A.J. Peer, Self and Tutor Assessment: Relative Reliabilities. *Stud. High. Educ.* **1994**, *19*, 69–75. [\[CrossRef\]](#)
74. Topping, K. Self and Peer Assessment in School and University: Reliability, Validity and Utility. In *Optimising New Modes of Assessment: In Search of Qualities and Standards*; Segers, M., Dochy, F., Cascallar, E., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2003; pp. 55–87. [\[CrossRef\]](#)
75. Blue, G.M. Self-Assessment—The Limits of Learner Independence. In *Individualization and Autonomy in Language Learning: ELT Documents 131*; Brookes, A., Grundy, P., Eds.; Modern English Publications: Hongkong, China, 1988; pp. 100–118.
76. Matsuno, S. Self-, Peer-, and Teacher-Assessments in Japanese University EFL Writing Classrooms. *Lang. Test.* **2009**, *26*, 75–100. [\[CrossRef\]](#)
77. Hawe, E.; Parr, J. Assessment for Learning in the Writing Classroom: An Incomplete Realisation. *Curric. J.* **2014**, *25*, 210–237. [\[CrossRef\]](#)
78. Bouziane, A.; Zyad, H. The Impact of Self and Peer Assessment on L2 Writing: The Case of Moodle Workshops. In *Assessing EFL Writing in the 21st Century Arab World: Revealing the Unknown*; Ahmed, A., Abouabdelkader, H., Eds.; Palgrave Macmillan: Cham, Switzerland, 2018; pp. 111–135. [\[CrossRef\]](#)
79. Butler, Y.G. Young Learners' Processes and Rationales for Responding to Self-Assessment Items: Case of Generic Can-Do and Five-Point, Likert-Type Formats. In *Useful Assessment and Evaluation in Language Education*; Davis, M.J., Norris, J.M., Malone, M.E., McKay, T.H., Son, Y.-A., Eds.; Georgetown University Press: Washington, DC, USA, 2018; pp. 21–40.
80. Yan, Z.; Brown, G.T.L. A Cyclical Self-Assessment Process: Towards a Model of How Students Engage in Self-Assessment. *Assess. Eval. High. Educ.* **2017**, *42*, 1247–1262. [\[CrossRef\]](#)



81. Gardner, D. Self-Assessment for Autonomous Language Learners. *Links Lett.* **2000**, *7*, 49–60.
82. Hobson, E.H. Encouraging Self-Assessment: Writing as Active Learning. *New Dir. Teach. Learn.* **1996**, *67*, 45–58. [\[CrossRef\]](#)
83. Lam, R. Assessment as Learning: Examining a Cycle of Teaching, Learning, and Assessment of Writing in the Portfolio-Based Classroom. *Stud. High. Educ.* **2016**, *41*, 1900–1917. [\[CrossRef\]](#)
84. Andrade, H.L.; Wang, X.; Du, Y.; Akawi, R.L. Rubric-Referenced Self-Assessment and Self-Efficacy for Writing. *J. Educ. Res.* **2009**, *102*, 287–302. [\[CrossRef\]](#)
85. Wang, W. Using Rubrics in Student Self-Assessment: Student Perceptions in the English as a Foreign Language Writing Context. *Assess. Eval. High. Educ.* **2016**, *42*, 1280–1292. [\[CrossRef\]](#)
86. Riazi, M.; Shi, L.; Haggerty, J. Analysis of the Empirical Research in the Journal of Second Language Writing at Its 25th Year (1992–2016). *J. Second. Lang. Writ.* **2018**, *41*, 41–54. [\[CrossRef\]](#)
87. Wanner, T.; Palmer, E. Formative Self-and Peer Assessment for Improved Student Learning: The Crucial Factors of Design, Teacher Participation and Feedback. *Assess. Eval. High. Educ.* **2018**, *43*, 1032–1047. [\[CrossRef\]](#)
88. Birjandi, P.; Hadidi Tamjid, N. The Role of Self-, Peer and Teacher Assessment in Promoting Iranian EFL Learners' Writing Performance. *Assess. Eval. High. Educ.* **2012**, *37*, 513–533. [\[CrossRef\]](#)
89. Liu, H.; Brantmeier, C. "I Know English": Self-Assessment of Foreign Language Reading and Writing Abilities among Young Chinese Learners of English. *System* **2019**, *80*, 60–72. [\[CrossRef\]](#)
90. Birjandi, P.; Siyyari, M. Self-Assessment and Peer-Assessment: A Comparative Study of Their Effect on Writing Performance and Rating Accuracy. *Iranian J. Appl. Linguist.* **2010**, *13*, 23–45.
91. Mok, M.M.C.; Lung, C.L.; Cheng, D.P.W.; Cheung, R.H.P.; Ng, M.L. Self-Assessment in Higher Education: Experience in Using a Metacognitive Approach in Five Case Studies. *Assess. Eval. High. Educ.* **2006**, *31*, 415–433. [\[CrossRef\]](#)
92. Bai, B.; Chao, G.C.N.; Wang, C. The Relationship between Social Support, Self-Efficacy, and English Language Learning Achievement in Hong Kong. *TESOL Q.* **2019**, *53*, 208–221. [\[CrossRef\]](#)
93. Lee, I.; Coniam, D. Introducing Assessment for Learning for EFL Writing in an Assessment of Learning Examination-Driven System in Hong Kong. *J. Second. Lang. Writ.* **2013**, *22*, 34–50. [\[CrossRef\]](#)
94. Yan, Q.; Cheng, X.; Zhang, L.J. Implementing Classroombased Assessment for Young EFL Learners in the Chinese Context: A Case Study. *Asia Pac. Educ.* **2021**, *30*, 541–552. [\[CrossRef\]](#)
95. Liu, J. Self-Assessment of English Writing Skills by Chinese University Students. *Mod. Foreign Lang. Q.* **2002**, *25*, 241–249.
96. Zheng, H.; Huang, J.; Chen, Y. Effects of Self-Assessment Training on Chinese Students' Performance on College English Writing Tests. *Polyglossia* **2007**, *23*, 33–42.
97. Ding, Y.; Zhao, T. Chinese University EFL Teachers' and Students' Beliefs about EFL Writing: Differences, Influences, and Pedagogical Implications. *Chin. J. Appl. Linguist.* **2019**, *42*, 163–181. [\[CrossRef\]](#)
98. Li, H.H.; Zhang, L.J.; Parr, J.M. Small-Group Student Talk before Individual Writing in Tertiary English Writing Classrooms in China: Nature and Insights. *Front. Psychol.* **2020**, *11*, 570565. [\[CrossRef\]](#)
99. Glasswell, K.; Parr, J.; Aikman, M. Development of the AsTTle Writing Assessment Rubrics for Scoring Extended Writing Tasks. *Univ. Auckland. Asttle Proj.* **2001**, *6*, 1–27.
100. IELTS; British Council. IELTS Task 2 Writing Band Descriptors (Public Version). 2015. Available online: [https://takeielts.britishcouncil.org/sites/default/files/IELTS\\_task\\_2\\_Writing\\_band\\_descriptors](https://takeielts.britishcouncil.org/sites/default/files/IELTS_task_2_Writing_band_descriptors) (accessed on 2 September 2022).
101. Jacobs, H.; Zingraf, S.; Wormuth, D.; Hartfiel, V.; Hughey, J. *Testing ESL Composition: A Practical Approach*; Newbury House: Rowley, MA, USA, 1981.
102. Smagorinsky, P. The Method Section as Conceptual Epicenter in Constructing Social Science Research Reports. *Writ. Commun.* **2008**, *25*, 389–411. [\[CrossRef\]](#)
103. Powers, D.E.; Fowles, M.E.; Farnum, M.; Ramsey, P. Will They Think Less of My Handwritten Essay If Others Word Process Theirs? Effects on Essay Scores of Intermingling Handwritten and Word-Processed Essays. *J. Educ. Meas.* **1994**, *31*, 220–233. [\[CrossRef\]](#)
104. Field, A.P. *Discovering Statistics Using SPSS*, 4th ed.; Sage: Thousand Oaks, CA, USA, 2013.
105. Cohen, J. *Statistical Power Analysis for the Behavioural Sciences*, 2nd ed.; Erlbaum: Hillsdale, NJ, USA, 1988.
106. Cohen, J. A Power Primer. *Psychol. Bull.* **1992**, *112*, 155–159. [\[CrossRef\]](#)
107. Raykov, T.; Marcoulides, G.A. *An Introduction to Applied Multivariate Analysis*; Routledge: New York, NY, USA, 2008.
108. Casal, J.E.; Lee, J.J. Syntactic Complexity and Writing Quality in Assessed First-Year L2 Writing. *J. Second. Lang. Writ.* **2019**, *44*, 51–62. [\[CrossRef\]](#)
109. Foster, P.; Skehan, P. The Influence of Planning and Task Type on Second Language Performance. *Stud. Second. Lang. Acquis.* **1996**, *18*, 299–323. [\[CrossRef\]](#)
110. Pajares, F.; Urdan, T. *Self-Efficacy Beliefs of Adolescents*; IAP: Greenwich, UK, 2005.
111. Hanrahan, S.J.; Isaacs, G. Assessing Self- and Peer-Assessment: The Students' Views. *High. Educ. Res. Dev.* **2001**, *20*, 53–70. [\[CrossRef\]](#)
112. Hattie, J.; Helen, T. The Power of Feedback. *Rev. Educ. Res.* **2007**, *77*, 81–112. [\[CrossRef\]](#)
113. Panadero, E.; Jonsson, A. The Use of Scoring Rubrics for Formative Assessment Purposes Revisited: A Review. *Educ. Res. Rev.* **2013**, *9*, 129–144. [\[CrossRef\]](#)

114. Panadero, E.; Brown, G.; Courtney, M. Teachers' Reasons for Using Self-Assessment: A Survey Self-Report of Spanish Teachers. *Assess. Educ. Princ. Policy Pract.* **2014**, *21*, 365–383. [[CrossRef](#)]
115. Laveault, D.; Allal, L. Implementing Assessment for Learning: Theoretical and Practical Issues. In *Assessment for Learning: Meeting the Challenge of Implementation*; Laveault, D., Allal, L., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 1–16.
116. Meihami, H.; Razmjoo, S.A. An emic perspective toward challenges and solutions of self- and peer-assessment in writing courses. *Asian-Pac. J. Second. Foreign Lang. Educ.* **2016**, *1*, 9. [[CrossRef](#)]
117. Zhang, L.J.; Cheng, X.L. Examining the effects of comprehensive written corrective feedback on L2 EAP students' performance: A mixed-methods study. *J. Engl. Acad. Purp.* **2021**, *54*, 101043. [[CrossRef](#)]
118. Zhao, H.; Zhang, L.J. *Teaching Writing in English as a Foreign Language: Teachers' Cognition Formation and Reformation*; Springer Nature International: Cham, Switzerland, 2022. [[CrossRef](#)]
119. Wu, X.M.; Zhang, L.J.; Dixon, H.R. EFL teachers' understanding of assessment for learning (AfL) and the potential challenges for its implementation in Chinese university EFL classes. *System* **2021**, *101*, 102589. [[CrossRef](#)]
120. Wu, X.M.; Dixon, H.R.; Zhang, L.J. Sustainable development of students' learning capabilities: The case of university students' attitudes towards teachers, peers, and themselves as oral feedback sources in learning English. *Sustainability* **2021**, *13*, 5211. [[CrossRef](#)]
121. Al-Adwan, A.S.; Nofal, M.; Akram, H.; Albelbisi, N.A.; Al-Okaily, M. Towards a sustainable adoption of E-learning systems: The role of self-directed learning. *J. Inf. Technol. Educ. Res.* **2022**, *21*, 245–267. [[CrossRef](#)]
122. Al-Adwan, A.S.; Albelbisi, N.A.; Hujran, O.; Al-Rahmi, W.M.; Alkhalifah, A. Developing a holistic success model for sustainable e-learning: A structural equation modeling approach. *Sustainability* **2021**, *13*, 9453. [[CrossRef](#)]
123. Cheng, X.L.; Zhang, L.J. Sustaining university English as a foreign language learners' writing performance through provision of comprehensive written corrective feedback. *Sustainability* **2021**, *13*, 8192. [[CrossRef](#)]