

Exploring the Effectiveness of Formative Assessment
in the Context of a Chinese University

Mo Guo

Department of Integrated Studies in Education

McGill University, Montréal

August 2017

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree
of Master of Arts in Second Language Education

Copyright © Mo Guo 2017

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

Abstract

This quantitative study investigates the effectiveness of self assessment (SA) and peer assessment (PA) in the form of written feedback in a Chinese university context. As widely-advocated assessment approaches for formative purposes (FA), SA and PA have been extensively examined in different contexts; however, the combined effectiveness of SA and PA is under-explored. This study was conducted in Chinese EFL writing classrooms at the tertiary level with English major undergraduate students (N= 79). A three-session research process is designed which includes one training session, one essay writing session and one assessment session. In order to compare the effectiveness of SA, PA, and combined SA and PA, the amount of correct identified errors given by learners in each treatment condition was compared to that given by a highly proficient reviewer. The type of feedback offered in each condition was also examined. The results indicate that the combination of SA and PA does not generate more feedback than SA and PA alone. Also, the focuses of feedback generated through SA and PA are not significantly different. As a research projected situated in a special context, the implications and limitations are discussed with references from previous literature. Lastly directions for future research are suggested.

Keywords: formative assessment, self-assessment, peer assessment, feedback

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

Résumé

Cette étude quantitative s'intéresse à l'efficacité de l'auto-évaluation (SA) et de l'évaluation par les pairs (PA) sous forme de rétroactions écrites dans le contexte d'une université chinoise.

Types d'évaluations formatives (FA) largement préconisées, la SA et la PA ont déjà été scrutées de manière extensive dans maints contextes; cela dit, l'effet combiné de la SA et de la PA demeure peu fouillé. Cette étude a été menée dans une classe chinoise d'écriture EFL de niveau tertiaire avec des étudiants de premier cycle (N=79) en langue anglaise. Un processus de recherche sur trois sessions a été élaboré: une première sessions consacrée à la formation, une deuxième à la rédaction d'un essai et une troisième, à l'évaluation. Afin de comparer l'efficacité de la SA de la PA séparées et combinées, le nombre d'erreurs pertinentes identifiées par les apprenants dans chaque condition de traitement est comparé à celui déterminé par un réviseur hautement compétent. Le type de rétroactions générées dans chaque condition de traitement est aussi examiné. Les résultats démontrent que le fait de combiner SA et PA ne génère pas plus de rétroactions que de les considérer séparément. De plus, les centres d'attention des rétroactions générés par la SA et la PA ne comportent pas de différences significatives. En tant que projet de recherche mené dans un contexte spécifique, les implications et les limitations sont soutenues par des références aux travaux passés. Finalement, des pistes pour de futures recherches sont suggérées.

Mots-clés: évaluation formative, auto-évaluation, évaluation par les pairs, rétroaction

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

Acknowledgements

This thesis would not exist without the generous guidance and encouragement from my supervisor, Dr. Susan Ballinger. It was my great pleasure and honor to work under her supervision during the past two years. Her constructive criticism and professional dedication constantly inspires me during the whole writing process.

I would also like to express my gratitude to Dr. Carolyn Turner, who led me into the field of second language assessment and offered valuable feedback on the initial proposal. Her enthusiastic guidance and profound understanding of second language assessment contributed enormously in the construction of this research project. I am also very grateful to Dr. Roy Lyster, Dr. Mela Sarkar, Dr. Jesús Izquierdo, Dr. Ronald Morris and Dr. Dale Boyle for expanding my view on the complex yet fascinating topic of second language acquisition.

A special thank you needs to go to Dr. Jose Correa for his insightful take on statistical analysis. His wise words greatly enhanced the robustness of the data processing in the study. I sincerely appreciate his help along the way.

Furthermore, I would like to thank all the participants in the study. It was a delight to work with all of them. Their positive attitudes towards participating in this research project have made my work so effortless. I especially would like to thank the participating teachers for squeezing out time from their schedule to coordinate with me.

I would also like to thank all the friends I have met since I came to Montreal two years ago. With all of you, I have established a deep bond with this amazing city. Each day, I feel incredibly lucky to dwell here as a student.

Last but not least, I would like to thank my parents for financially and emotionally supporting my graduate studies. This journey would not start without their confidence in me.

Table of Contents

LIST OF TABLES	VIII
LIST OF FIGURES	IX
CHAPTER 1: INTRODUCTION.....	1
1.1 BACKGROUND	1
1.2 FORMATIVE ASSESSMENT (FA).....	1
1.3 PURPOSE STATEMENT	3
CHAPTER 2: LITERATURE REVIEW	5
2.1 INTRODUCTION.....	5
2.2 SITUATING FA IN SECOND LANGUAGE ASSESSMENT	6
2.3 SELF-ASSESSMENT (SA)	15
2.4 PEER ASSESSMENT (PA)	21
2.5 FEEDBACK IN FA	29
2.6 THE PRESENT STUDY	31
2.7 SUMMARY	32
CHAPTER 3: METHODOLOGY	34
3.1 INTRODUCTION.....	34
3.2 RESEARCH QUESTIONS.....	34
3.3 RESEARCH SETTING	35
3.4 PARTICIPANTS	38
3.5 ETHICAL CONSIDERATIONS	40
3.6 RESEARCH DESIGN.....	41
3.7 DATA COLLECTION INSTRUMENTS	46

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

3.8 DATA ANALYSIS	49
3.9 SUMMARY	49
CHAPTER 4: RESULTS	50
4.1 INTRODUCTION.....	50
4.2 RQ1: DESCRIPTIVE RESULTS	51
4.3 RQ1: STATISTICAL ANALYSIS.....	56
4.4 RQ2: DESCRIPTIVE RESULTS	59
4.5 RQ2: STATISTICAL ANALYSIS.....	63
4.6 SUMMARY	66
CHAPTER 5: DISCUSSION	67
5.1 INTRODUCTION.....	67
5.2 FINDINGS	67
5.3 IMPLICATIONS	73
5.4 LIMITATIONS.....	77
5.5 DIRECTIONS FOR FUTURE RESEARCH.....	80
5.6 SUMMARY	82
CHAPTER 6. CONCLUSION.....	83
REFERENCE.....	85
APPENDIX A: INITIAL COMMUNICATION.....	99
APPENDIX B: STUDENT CONSENT FORM.....	101
APPENDIX C: TEACHER CONSENT FORM.....	104
APPENDIX D: ANALYTIC RUBRICS	107
APPENDIX E: TEM4 WRITING PROMPT	109

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

APPENDIX F: REFLECTIVE SHEET.....	110
-----------------------------------	-----

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

List of Tables

Table 1	Test content and format of TEM4
Table 2	Demographic information of student participants
Table 3	Student participants group plan
Table 4	descriptive presentation of SCIE-OE discrepancy
Table 5	descriptive presentation of SCIE/OE percentages by error type
Table 6	descriptive presentation of SCIE-OE discrepancy
Table 7	Test of homogeneity of variance results
Table 8	One-way ANOVA test results
Table 9	Descriptive presentation of the frequency counts of SCIE in SA and PA
Table 10	descriptive presentation of SCIE in each error type of written work/ SCIE total
Table 11	descriptive presentation of SCIE/OE total
Table 12	Mean ranks from Mann-Whitney U test
Table 13	Mann-Whitney U test results

List of Figures

- | | |
|----------|---|
| Figure 1 | Agents of learning-oriented assessment |
| Figure 2 | Self-assessment in self-regulation of learning |
| Figure 3 | TEM4 passing rates comparison (2011- 2016) |
| Figure 4 | Three-step process of the second research session |
| Figure 5 | Process of data collecting and coding |
| Figure 6 | Procedures of SA (left) and PA (right) |
| Figure 7 | Split histograms of SCIE-OE discrepancy |
| Figure 8 | Split histograms of SCIE/OE total |

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

Exploring the Effectiveness of Formative Assessment in the Context of a Chinese University

Chapter 1: Introduction

1.1 Background

Since the late 1950s, English has been the most important foreign language in China (Lam, 2005), and the pedagogy of English as a foreign language (EFL) has witnessed a long process of development over the past decades. However, in English writing classrooms at Chinese universities, Confucianism remains the core cultural origin of all pedagogical practices, making these classroom contexts distinct from that of western universities (Jiang, 2013). Confucianism is a philosophical belief that prioritizes “gaining wisdom and complying with morality” as well as the formation of “group identity and harmony maintenance” (Jiang, 2013, p. 96). When this philosophy is applied to the EFL classrooms, it shapes the classroom into a highly teacher-centered learning environment. As pointed out by Jiang (2013), the general perception of a typical Chinese classroom is that “what teachers say is authoritative and true and these truths should be repeated and remembered instead of being questioned by learners” (p. 96).

Tertiary-level EFL instruction in China, in particular, adopts a “teacher-centered textbook-analysis-based Grammar-Translation method” (Yang, 2000, p.19). It attaches great importance to grammatical accuracy and vocabulary abundance, especially as a preparation for high-stakes tests which is regarded as a central part of EFL in Chinese universities (Mattisson Ekstam, 2015).

1.2 Formative Assessment (FA)

As opposed to summative assessment, FA takes place within instruction sessions to promote learners’ knowledge acquisition and to facilitate instructors’ pedagogical decision-

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

making. It provides students with access to constructive feedback that can help them progress academically. According to Cizek (2010), there are two criteria to be met for an assessment to be labelled as formative. Firstly, FA is conducted during course time. Secondly, the main aim of FA should be one of the following:

- to shed light on what the students have or have not learned;
- to help teachers gather information for future lesson planning;
- to assist students in evaluating their academic progress, to improve their own work, and to learn how to self-assess and to self-regulate learning.

In other words, FA serves as a complementary part of instruction and benefits both instructors and learners. In recent years, FA has gained more attention from academia. Cizek (2010) claimed that the reason for this attention is education stakeholders realise that FA offers valuable information on students' academic achievements that other assessment forms cannot. Therefore, it represents "...the next best hope for stimulating gains in student achievement" (Cizek, 2010, p. 3).

In practice, FA takes various forms. For instance, with the involvement of students as assessors, there is SA and PA. The effectiveness of SA and PA in improving second language (L2) learners' writing competence has been investigated in a large variety of contexts. For instance, Strong-Crause (2000) pointed out that for assessing L2 writing skills, SA is an effective tool as students are capable of making a reliable and valid assessment of their own writing if detailed instructions are provided. Also, Anderson (2012) claimed that students could benefit from SA regardless of their academic standing. Topping (2009) exhaustively concluded the benefits of PA. According to him, successfully implemented peer assessment can give constructive feedback to students; the gains from peer assessment are beneficial to both the

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

assessors and the assessed; the positive results of peer assessment in writing are remarkable, in particular with the implementation of peer editing; it promotes cooperative learning as it involves communication and negotiation among peers; it also saves teachers' time spent on the rating process.

Building on previous literature, the present study investigates the effectiveness of FA in the context of a Chinese university. In China, FA has been promoted through institutional documents, however, because of the overwhelming focus on high-stakes tests, the implementation of classroom-based FA in actual EFL writing classrooms is rare (Chen, May, Klenowski & Kettle, 2014). Therefore, Chinese university students, such as those who participated in the present study, usually have very limited knowledge of or experience with SA and PA.

1.3 Purpose Statement

This study is a quantitative analysis of feedback produced through SA and PA. As a branch of classroom-based assessment (CBA), the FA practices investigated in the present study fit in the research agenda proposed by Turner and Purpura (2016) and further the theoretical connection between FA and learners' independence and autonomy, which has been identified as one of the primary goals of higher education in 21st century (Germaine, Richards, Koeller & Schubert-Irastorza, 2016). In addition, the rationale of the study answers the call for more research studies on the formative purpose of language assessment (Way, Dolan & Nicolas 2010; Carless, 2012; Stoyanoff, 2012; Turner, 2013; Brown, Andrade & Chen 2015) as well as positioning feedback quality at the center of conceptualizing FA practices (Brown & Harris, 2013a; Topping, 2003; Turner and Purpura, 2016). The FA practices used in the present study are viewed as teaching tools with the potential to facilitate learning rather than as evaluation

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

tools for performance measurement; and the feedback quality refers to whether or not the feedback is correct. By comparing the feedback from SA and PA, this study also reveals the characteristics of feedback on students' written work, which highlights the differences between SA and PA. On a theoretical level, it serves as a comparison between the metacognitive function of self-reflection and peer interaction.

Mixed results from different research contexts have been shown about the reliability of SA and PA as performance measurement tools, which implies the significance of context in the implementation of FA. In addition, very few studies of SA and PA have taken place in Chinese universities. Through this research study, the challenges facing the implantation of FA in the special context of one Chinese university will be revealed.

Moreover, the robust formative function and students' positive attitudes towards SA and PA as shown in previous studies indicate that they have significant pedagogical potential. However, studies on SA and PA, to a great extent, focus on the discrepancy of scoring between teachers' ratings and students' ratings. There is a lack of attention to the actual feedback produced through the FA process (Stoynoff, 2012; Turner, 2013). Several studies focusing on feedback have reported very promising results about using the feedback to enhance the formative function of SA and PA (Cho & Cho, 2011; Nicol, Thomson & Breslin, 2014). Few studies have looked at the combined effects of SA and PA even though both are advocated individually. As two closely related assessment approaches with shared formative function (Carnell, 2016), there is an emerging need to study them as a dynamically synthesized combination.

Chapter 2: Literature Review

2.1 Introduction

Modern language assessment has been around since the 20th century (O'Sullivan, 2012). Nowadays, it is used as a general term for an operationalized process for implementing and interpreting both conventional and alternative tests to generate “inferences or claims about certain language-related characteristics of an individual” (Purpura, 2016, p.191). In other words, it covers both high-stakes tests external to the classroom such as the Test of English as a Foreign Language (TOEFL) administered by a private nonprofit educational testing organization, and CBA embedded in instruction sessions such as teachers' ongoing FA, students' SA and PA.

In the context of second language (L2) assessment, Bachman and Palmer (2010) identified three common attributes: first, assessments are designed upon substantial foundation, meaning the content of assessments derives from instruction, rubrics and psychometric measurement models. Second, regardless of the format of implementation, assessments share the goal of facilitating decision-making. Lastly, the assessments have some level of systematicity. Be it a timed test or a classroom routine, there are operationalized procedures for instructors and learners to follow. The purpose of conducting an L2 assessment, according to Purpura (2016), is to

elicit the L2 performance from an individual under certain conditions so that performance consistencies can be interpreted and used to produce records such as scores, verbal descriptions, or mental notes. Interpretations from these records are then used as evidence for making decisions (p. 191).

The broad range covered by the term L2 assessment has triggered an explosion of discussion in recent years regarding key definitions within the scope of L2 assessment. In this

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

chapter, I will review key definitions developed over these years in L2 assessment research for the purposes of delineating the topic and scope of the present research study. These include classroom-based assessment (learning-oriented assessment), assessment for learning (formative assessment), self-assessment and peer assessment. The effectiveness of these different forms of classroom-based assessments will be triangulated with a synthesis of relevant research papers. Also, as the primary parameter of the study design, feedback and its significance will be briefly discussed. Finally, this chapter will end with a discussion of the scholarly gap that this thesis attempts to fill.

2.2 Situating FA in Second Language Assessment

2.2.1 Classroom-based assessment (CBA). L2 assessment, in general, could be categorized into two categories: assessment outside the classroom and assessment inside the classroom (Turner & Purpura, 2016). Since the 1990s, the focus of research on L2 assessment has experienced a transition from assessment external to the classroom to assessment internal to the classroom. The term CBA has emerged during this time; however, since the description process is still in its infancy, there is no consensus on the definition of this term despite the fact that scholars have provided many interpretations of it. For instance, Leung (2004) defined CBA as “nonstandardized local assessment carried out by teachers in the classroom”, which indicates that the implementation of CBA is embedded in classroom pedagogical practices and the negotiation between students and teachers is an essential component in CBA. Another group of researchers, Hunter, Mayenga, and Gambell (2006) defined CBA as an interactive effort between students’ performance and teachers’ evaluation in the daily learning environment. These authors also considered factors related to learners’ attributes such as attitudes towards assessment practices. Hill and McNamara (2011), in response to the call for a separated diagram for CBA

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

(Leung & Lewkowicz, 2006; James, 2006; Davison & Leung, 2009), drew on research data from two foreign language learning classrooms in Australia and constructed an empirically-based research framework for CBA. Within their framework, they give the following definition of CBA: “Any reflection by teachers (and/or learners) on the qualities of a learner’s (or group of learners’) work and the use of that information by teachers (and or learners) for teaching, learning (feedback), reporting, management or socialization purposes.” More Recently, Turner (2013), drawing on the concepts in previous studies, provided the definition depicting key characteristics of CBA:

It involves strategies by teachers to plan and carry out the collection of multiple types of information concerning student language use, to analyze and interpret it, to provide feedback, and to use this information to help make decisions to enhance teaching and learning. Observable evidence of learning (or lack of learning) is collected through a variety of methods, and most often embedded in regular instructional activities. (p. 66)

These definitions are just several examples of the fact that the complexity of CBA keeps prompting researchers to elicit what this term entails exactly.

As a significant branch of CBA, Purpura (2004) coined the term Learning-Oriented Assessment (LOA) in a paper discussing the pedagogical practices in the evaluation of grammar. Since then, LOA approaches have started to emerge in research studies (Boud & Falchikov, 2006; Carless, 2007; Lombard, 2008; Carless, 2015; Vanderlelie and Alexander, 2016) featuring the localization of assessment. Carless (2007) identified three traits to outline the realm of LOA: students’ involvement in the decision-making process, the constructive nature of feedback from assessment, and assessment as part of instruction.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

In an attempt to further situate LOA within the research agenda of CBA, Turner and Purpura (2016) identified three key characteristics of LOA. First, feedback or other forms of complementary information takes the central role in L2 knowledge construction towards a certain learning objective. Second, LOA is predominantly an independent cognitive process. Meanwhile, it is also a “highly intricate socio-cognitive and sociocultural process” in a communicative learning environment (p. 261). Last, the nature of LOA in the L2 acquisition classroom is multifaceted. In other words, it is under the influence of many factors related to all stakeholders.

Turner and Purpura (2016) further identified seven interrelated dimensions (See Figure 1) to define LOA and, therefore, included LOA under the research agenda of CBA. These dimensions are: the contextual dimension (i.e. the sociocultural characteristics of teachers and students), the elicitation dimension (i.e. types of assessment practices chosen by teachers), the proficiency dimension (i.e. coordination of learners with different L2 proficiency), the cognitive or learning dimension (i.e. peer feedback and self-regulation), the affective dimension (i.e. attitudes and emotions related to assessment), the interactional dimension (i.e. communication taking place during the assessment process), and the instructional dimension (i.e. teachers’ L2 knowledge and content knowledge).

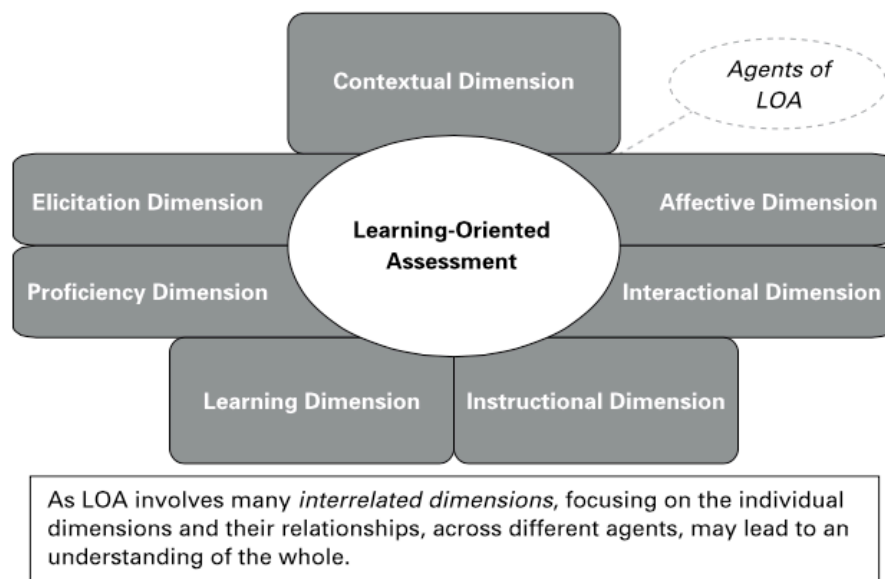


Figure 1. *Agents of learning-oriented assessment (Turner & Purpura, 2016, p. 261)*

Purpura (2016) notes that the frequency that LOA appeared at influential conferences and symposia has noticeably increased in recent years. Therefore, it seems that the significance of this topic has been increasingly acknowledged. The constant development of LOA has significant implications and contributions to the theorization of CBA.

One CBA approach, FA, will be discussed in detail in the following section.

2.2.2 Formative Assessment (FA). The origin of the term “formative assessment” can be traced back to 1967. It is believed that Michael Scriven (1967) coined the term by conceptualizing assessment processes that “have a role in the on-going improvement of the curriculum” (p.41). The focus of his definition is the value of FA at the program level. At the classroom level, Bloom (1969) proposed that FA be implemented as an “evaluation by brief tests used by teachers and students as aids in the learning process” (p. 48). Thereby, the scope expanded to the functions FA serves inside the classroom. However, Bloom’s definition was still isolated from classroom instruction; it merely referred to the use of formal assessment

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

procedures for the purpose of gathering information to facilitate teaching (Wiliam, 2014). Sadler (1989) claimed that FA should be entwined with sound instruction: “Formative assessment is concerned with how judgments about the quality of student responses (performances, pieces, or works) can be used to shape and improve the student’s competence by short-circuiting the randomness and inefficiency of trial-and-error learning (p. 120).” This reconceptualization of FA beyond formal tests is also supported by Torrance (1993) who suggested that classroom interactions should also be included in a fundamental review of L2 assessment.

Along with the evolution of the theoretical foundation of FA, the formation of an explicit definition has also experienced a long process of construction. In a substantial review of 250 book chapters and journal articles on FA practices in mathematics, science, and other subjects, Black and Wiliam (1998a) defined FA as:

All those activities undertaken by teachers—and by their students in assessing themselves—that provide information to be used as feedback to modify teaching and learning activities. Such assessment becomes formative assessment when the evidence is actually used to adapt the teaching to meet student needs (p. 140).

This definition clarifies that the word ‘formative’ is not an inherent attribute of any particular assessing practice but rather an adjective to describe how assessment processes influence learning and teaching. Afitska (2014) summarized four core characteristics of FA: Firstly, FA should be a complementary part of the instruction process that helps instructors gather information for future course planning. Secondly, the nature of FA should be learning-facilitating. Thirdly, both the educators and learners should benefit from the evaluation process. Lastly, FA should prompt a communicative classroom setting, either between teacher and students or students and students.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

Some scholars are not content with the ambiguity brought by the breadth of the term “formative” in categorizing the assessment processes with the above features and favored another term “assessment for learning” (AFL) coined by Harry Black (1986), which has been defined as “the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning where they need to go and how best to get there.” (Broadfoot, Daugherty, Gardner, Harlen, James, and Stobart, 2002).

However, as Bennett and Gitomer (2009) argued, the different wording solely contributes to clarifying definitional borders. Most researchers still use the terms formative assessment and assessment for learning interchangeably in their research (Afitska, 2014; Baleni, 2015; Chen, May, Klenowski & Kettle, 2014; Ketabi & Ketabi, 2014; Lau, 2016; Lee & Coniam, 2013; Purpura, 2016; Rashid & Jaidin, 2014;), but some scholars further distinguish these two terms due to their different scopes of coverage (e.g. Gardner, 2012; Wiliam, 2011). In this thesis, these two terms will be used as synonyms.

The fuzziness of the definition of FA demands further categorization of certain classroom assessment approaches to more precisely and properly situate the present research in the CBA research agenda proposed by Turner and Purpura (2016).

By the categorization of Yorke (2003), FA has two forms: formal and informal. Formal FA is, usually but not always, carried out “by academic staff or supervisor of a placement activity within a collaborating organization” (p. 479). It may engage students in the assessing process. For example, students compose midterm papers according to the syllabus, and then teachers provide corresponding feedback that students can learn from. What is worth noticing here is the formative use of feedback. If feedback is not available to students after the evaluation process, the assessment cannot be characterized as formative anymore. This corresponds to Lau

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

(2016) as she argues that PA and SA are two connected assessment conceptions and confirms that whether assessments are formative or summative usually depends on implementation, execution, and interpretation (Lam, 2013; Sambell, McDowell & Montgomery, 2012).

Informal FA usually happens during the class instruction time but not because of curriculum regulation (Yorke, 2003). It covers a wide range of cognitive activities, such as instructors giving feedback on students' learning activity instantaneously and commenting on draft compositions of students for further improvement. As argued by Yorke (2003), FA could be "very occasional", but it should essentially be supportive of knowledge acquisition (p. 479). FA could be implemented by anyone. Learners may be able to receive feedback from assessors external to the classroom, such as peers from different disciplines. Informal FA can happen indirectly when learners reflect the feedback they obtain from reviewing peers' work on their own work.

By this categorization, the assessment approaches used for formative purposes in the present research, SA and PA, are informal FA. They are treated as an integral part of the battery of pedagogical practices with the main purpose of facilitating the learning process.

There are a substantial quantity of research papers examining the benefits of FA. According to Song (2011), three themes emerge from a systematic review of past literature:

- FA facilitates knowledge acquisition;
- FA offers reciprocal feedback to both instructors and learners;
- FA promotes learners' engagement in learning.

Each theme will be discussed in detail respectively in the following section.

Firstly, FA facilitates knowledge acquisition. Anderson and Palm (2017) reported that after a ten-month exposure to FA, students witnessed significantly higher academic progress than

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

student who did not participated in FA. Similarly, Black and Wiliam (1998b) claimed that FA is quantitatively effective in improving academic performance, and especially beneficial for students with poor academic performance. Moreover, they pointed out the crucial role of feedback in guiding instruction adjustments. Interestingly, they admitted later in another paper that empirical judgements were applied in interpreting their previous research results and therefore, reached an overly optimistic conclusion that there is sufficient evidence to triangulate the effectiveness of formative assessment in the classroom (Black & Wiliam, 2003). However, empirical experience is a valuable resource when it comes to research in educational settings (Cohen & Morrison, 2013). The reference value of their study is still relatively high.

Secondly, FA offers reciprocal feedback to both instructors and learners. Here, feedback has a different meaning compared to feedback in second language acquisition. It refers to “the information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way” (Afitska, 2014). It serves as a cognitive bridge for learners to move from where they are to where they wish to be in the learning process. Feedback, in this sense, is a fundamental feature of FA.

For instructors, Ertle, Rosenfeld, Presser and Goldstein (2016) argued that feedback from FA deepens instructors’ understanding of the teaching materials and the cognitive process of learning. With sufficient professional development training, FA feedback should help teachers to “translate all that information into better instruction” (p.987). For learners, Graham, Hubert and Harris (2015), in a meta-analysis of true and quasi-experimental research studies about FA in writing classrooms, found that FA feedback is not only beneficial for the improvement of students’ written work but also for the acquisition of writing strategies and skills. Moreover, their

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

findings revealed that FA feedback benefits learners regardless of the variety of sources, such as instructors, peers or self.

In addition, Kulik and Kulik (1988) argued that feedback, produced during and after the evaluation process, is equally valuable. Specifically, they found that immediate feedback contributed to the flow of the evaluation process while delayed feedback contributed to the overall task performance.

According to Tan (2013), whether FA feedback fulfills its potential to facilitate instructors' decision-making and to inform learners of their knowledge gap depends on two conditions: the length of assessment training and "the realistic gap in students' learning that may be bridged within that time" (p. 3). He claimed that these two conditions determine whether the positive impact of feedback extend beyond the "immediate/imminent context" (Tan, 2013, p. 3). In other words, the effects of feedback in FA are under the restriction of exposure to assessment and the level of initial academic knowledge of learners.

The significance of feedback does not stop here. As a central part of FA practices, it will be elaborated in the later section.

Thirdly, FA promotes learners' engagement in learning. Feedback serves multiple functions in coordinating students' learning. Except for contributing to immediate quantifiable improvements in students' academic performance, it also motivates students to pursue further learning (Heritage, 2007). It ensures that strengths of students' academic performance are recognized and appreciated, and that gaps between the present status and the expected status are identified and filled (Fishbach, Eyal, & Finkelstein, 2010). McMillan, Cohen, Abrams, Cauley, Pannozzo and Hearn (2010) found that there was a statistically significant positive relationship between FA and student motivation; and students reported higher levels of motivation when

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

diversified FA practices were introduced to the classroom. Similarly, Weurlander, Soderberg, Scheja, Hult and Wernerson (2012) claimed that FA can be viewed as “a tool for learning, contributing to the process and outcomes of learning” (p. 747) as learners felt motivated by the implementation of FA practices.

As noted by Wiliam (2013), the rationale supporting that FA promotes students’ engagement in learning is “when students come to believe that smart is not something you are but something you get, they seek challenging work, and in the face of failure, they increase effort” (p. 13).

The assessment approaches used for formative purposes in the current study, SA and PA, have been increasingly applied in L2 education in recent years. The following two sections will be dedicated to reviewing the concepts and effectiveness of these two practices respectively.

2.3 Self-Assessment (SA)

In this section, I will present the rationale of SA, including its definition, features, and benefits, then provide evidence from previous literature supporting the effectiveness of SA.

2.3.1 Rationale of SA. As an assessment approach widely used for formative purposes during classroom teaching, the definition and design of SA vary greatly across different contexts. For the present study, the definition proposed by Andrade and Valtcheva (2009) is adopted. They define SA as “a process during which students collect information about their performance or progress; compare it to explicitly stated criteria, goals, or standards; and revise accordingly” (p. 13). Their definition highlighted the formative function of this evaluation approach and emphasized that SA is not about assigning a grade. The main reasons for implementing SA among learners are “to boost learning and achievement and to promote academic self-regulation, or the tendency to monitor and manage one’s learning” (Andrade & Valtcheva, 2009, p. 13).

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

This dynamic relationship between formative SA and self-regulation of learning is best demonstrated by the following Figure 2 from Andrade (2010):

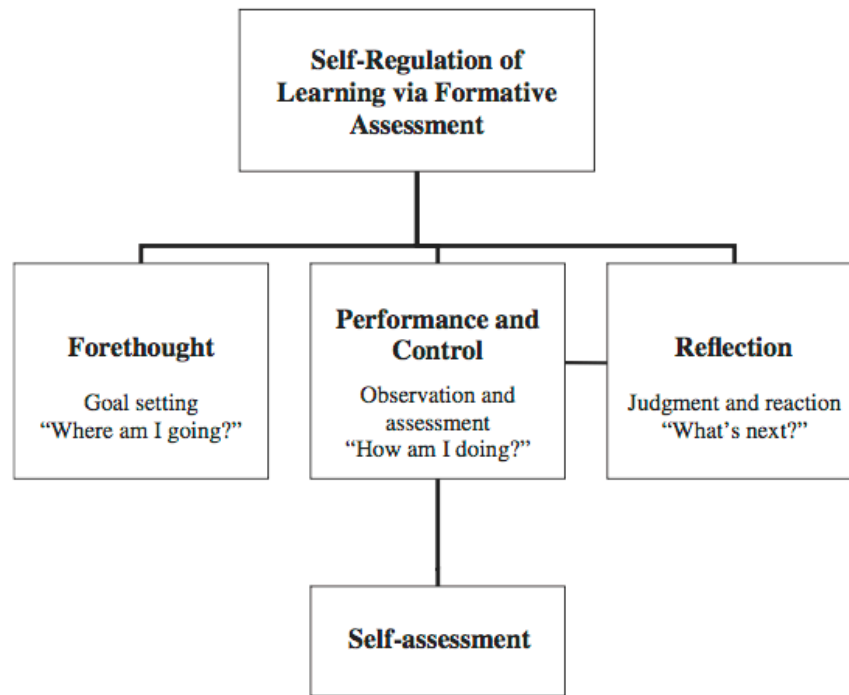


Figure 2. *Self-assessment in self-regulation of learning (Andrade, 2010, p. 95)*

Andrade (2010) identified three core features of effective SA: first, elicited rubrics are available to student assessors. There might be a negotiation between teachers and students for clarification before the assessment process begins. One way to achieve common understandings of rating rubrics is to guide students through the practice while using models. Second, critical reviewing is conducted by students on their own work referring to the rating rubrics. Third, revision is implemented on the draft work with feedback from the previous reviewing. It provides the opportunity for students to gain explicit improvement. Admittedly, there are other factors influencing the SA process (Goodrich, 1996); these are only the fundamental features that apply to the majority of SA practices.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

The majority of research on SA still concentrates on comparing the convergence between learners' quantitative evaluation results to those of teachers. In other words, the effectiveness of SA is determined by the summative grading procedures, not the provision of formative feedback. Brown, Andrade, and Chen (2015) called for a shift of focus in research from grading to "formative feedback" (p. 444). They claimed that due to the unique function that SA serves, it is reasonable to reconsider using only quantitative evaluation results to determine accuracy in the context of SA. SA has the potential to achieve this goal because it is an easily accessible form of feedback on learners' academic competence and products. Similarly, Brown and Harris (2014) argued that SA should "no longer be treated as assessment, but instead as an essential competence of self-regulation" (p. 22).

2.3.2 Effectiveness of SA. The effectiveness of SA has been extensively tested in various contexts. A ground-breaking literature review conducted by Falchikov and Boud (1989) of 48 quantitative studies from the 20th century found that most studies yielded positive effect sizes (between - 0.62 and 1.42 with the mean of 0.47) and demonstrated fairly strong correlation coefficients (r) between teacher and student scores (between -0.05 and 0.82 with the mean being 0.39) (pp. 419-420). Later, Brown and Harris (2013a) narrowed the mean level of agreement between SA and external measurement of performance to $r=0.3$ to 0.5 .

Falchiko and Boud (1989) also identified three factors to explain the variance in findings on SA: a) the methodological quality of the research; b) the level of difficulty of the course that participants were enrolled in; and c) the nature of the field of knowledge. Specifically, more rigorously designed research, learners enrolled in courses at a higher academic level (i.e. graduate level versus undergraduate level), and science-oriented disciplines generate a stronger effect size for SA. Panadero, Brown and Strijbos (2016) added another two factors contributing

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

to the variance, which were students' involvement in the development of rating rubrics and the availability of "concrete, specific and well-understood criteria or reference points" (p. 815).

Contemporary studies on second language assessment further contribute to the theoretical and empirical strength of SA. In North America, Dolosic, Brantmeier, Strube and Hogrebe (2016) examined how SA contributed to 24 French learners' oral communication competence. They administered students' SA through a criterion-referenced questionnaire during a French language summer camp session. Their findings revealed that SA has a strong pedagogical potential to serve as a self-diagnosis tool in classroom settings, and with access to the evaluation results, teachers can better respond to individual differences in L2 acquisition. Moreover, the researchers also pointed out that teacher's mediation and real-time instruction during the evaluation process are beneficial for the effective implementation of SA.

Similarly, Lappin-Fortin and Rye (2014), using a pre-posttest design, investigated the divergence between SA and teacher assessment (TA) in a university-level French pronunciation course for L2 learners. They found that some aspects of students' oral competence witnessed more prominent progress from SA. To be specific, learners' general assessment and certain aspects of pronunciation were significantly improved, especially when there were precise rules of evaluation involved. Besides the pronunciation, learners also became more aware of the various types of errors, which indicates the great pedagogical potential of SA. Mahlberg (2015) conducted SA on first- and second-year community college students. Ten SA and six traditional assessment classes completed the Motivated Strategies for Learning Questionnaire at the end of one semester. The results indicated that classes with SA demonstrated higher self-regulation and a significant increase in students' intention to pursue further study.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

In Europe, Suñol, Arbat, Pujol, Feliu, Fraguell & Planas-Lladó (2016) conducted their research study with participants from a university in Spain enrolled in 7 different degree subjects and fields of knowledge. In order to investigate the possibility of using SA as a complementary part of assessment conducted by the instructor, the researchers compared the results from SA and TA. Given the high deviation between these two assessment approaches, the authors claimed that the formative function of SA is more valuable than the summative counterpart.

In Oceania, Boud, Lawson, and Thompson (2015) looked into the development process of learners' competence in performing SA. Participants from two Australian universities were asked to conduct SA on their written work on a regular and voluntary basis for one academic year. Then the researchers compared students' criteria-based scorings to those assigned by instructors. The results indicated that participants' ability of conducting SA is cultivatable through repeated exposure to the evaluation process, which emphasizes the significance of including SA into daily instruction for the purpose of maximizing its effectiveness. Also, they found that the diversity of assessment approaches inside the classroom boosts learning efficiency; therefore, they advocated instructors including multiple evaluation patterns in curriculum design.

In Asia, Baleghizadeh and Hajizadeh (2014) investigated the development of the SA ability of 15 Iranian EFL learners. They exposed participants to instructor-assigned scores after each SA session, and by the end of the fourth week, they found that there was a high correlation between SA and TA in the evaluation of different components of academic writing. Based on the reading of data analysis results, they claimed that SA should not only be used as an assessment tool but also a powerful approach to cultivating competent writers.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

Similarly, Kilic (2016) compared the scores generated from SA and TA in teacher training at the tertiary level using a criteria-based assessment form. The participants evaluated their performance on oral presentations. The results indicated that results from SA and TA are closely related. The formative use of SA leads to better performance, improved self-confidence, and stronger presenting skills.

Suzuki (2014) conducted a research study on SA with adult advanced L2 learners of Japanese whose mother tongue is Chinese in a Japanese university. Learners were required to conduct SA on their acquisition of Japanese. She found that less experienced students tended to overestimate their competence, while more experienced learners tended to underestimate their proficiency. Students' judgment was less accurate when more challenging tasks were introduced into the classroom. This study confirms that experiential factors should be considered when the effectiveness of SA is discussed.

Similar results were reported by Ünalı (2016) as he compared the ratings from SA and TA on an objective placement test for intermediate level Iranian EFL learners. Also, the results from criterion-referenced evaluation process indicate that SA and TA were correlated. However, TA is a better predictor of learners' actual language proficiency. Yoon and Lee (2013) investigated the influence of SA on three groups of Korean EFL learners with mixed initial English proficiency. The participants in their study were 184 sixth-grade students from seven classes with low to high L2 proficiency level. The researchers instructed a SA session at the end of each class over 11 weeks. The pre-post test results indicate that SA positively influenced the L2 learning progress and learning motivation of students from all levels. Students with low and intermediate language proficiency levels benefited more from SA than that of students with high initial English proficiency.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

Shifting from the quantitative focus in the research of SA, Huang (2016) looked into the quality of feedback produced by SA. Participants were asked to listen, transcribe and analyze their speaking samples from the audio files of a foreign language speaking test as well as to propose suggestions for further improvement. As the author described, the self-feedback was “far-reaching and multifaceted” (p. 803), meaning it went beyond the students’ current learning stage and pointed to higher learning objectives. It revealed the great pedagogical potential of SA, and further attested that SA is more than just an evaluation tool, but rather an educational approach that requires further investigation.

2.4 Peer Assessment (PA)

In this section, I will present the rationale of PA, including its origin, definition, and benefits, then provide evidence from previous literature supporting the effectiveness of PA.

2.4.1 Rationale for PA. Peer assessment has been practiced in the classroom for over two centuries. As early as 1774, a British professor at the University of Glasgow proposed a curriculum including a detailed description of PA (Topping, 2003). About the action models, PA appears to be “very various in its implementation (Topping, 2010a, p. 339). As elicited by Topping (2003), PA could be conducted on “writing, portfolios, oral presentations, test performance or skilled behaviors” (p. 65). It could serve a summative or formative purpose. The curriculum areas, participants’ constellation, directionality, methods, and settings are of great diversity in practice.

Taking the multiple variables into consideration, PA has been defined as “an arrangement for learners and/or workers to consider and specify the level, value, or quality of a product or performance of other equal-status learners and/or workers” (Topping, 2003, p. 65). Topping (2009) further proposed a formative view of PA, under which the purpose of PA “is to help

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

students help each other plan their learning, identify their strengths and weaknesses, target areas for remedial action, and development metacognitive and other personal and professional skills” (p. 20). This view emphasizes the fruitfulness of PA as a way to produce constructive feedback (Topping, 2010b). The present study also adopts this formative view of PA.

Topping (2009) organized the benefits of formative PA into five broad categories: firstly, successfully implemented PA can provide constructive feedback to students. As long as the feedback is received with an open and positive attitude, it can contribute to the reduction of various types of errors and the development of self-regulation. This corresponds to Black and Wiliam’s (1998a) claim that regular constructive personalized feedback results in noticeable academic progress. Secondly, PA is beneficial for both the assessors and the assessed. Students as assessors can learn how to accurately interpret the rating rubrics; while the assessed can learn from the peer feedback as well as cultivate a stronger sense of self-reflection. This statement is consistent with Allen and Mills (2014) as they claim that learning through assessment is a reciprocal process for assessors and those being assessed. Thirdly, Topping (2009) cites the work of O’Donnell and Topping (1998) which indicates that the positive results of PA in writing are remarkable, in particular with the implementation of peer editing. However, a number of studies conducted on language competence components other than writing prove that peer assessment is an effective way of learning in L2 acquisition. Fourthly, PA promotes cooperative learning as it involves communication and negotiation among peers. This argument is supported by the expanding body of research on peer interaction in recent years. For example, Sato and Ballinger (2016) claimed that peer interaction, to some extent, is more beneficial than instructor-student interaction as a larger quantity of personalized feedback is available to learners. There is also evidence to testify that more chances for constructive peer interaction lead to more learning gains

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

(Netten & Spain, 1989). Lastly, with sufficient instruction and opportunities to practice, the reliability of PA will witness noticeable improvement. Therefore, it is a possibility for peer assessment to save the time teachers spend on the rating process. In addition to these five categories, PA is also believed to benefit teachers in the sense that it generates more precise and personalized pictures of students' academic performance (Van Zundert, Sluijsmans & Van Merriënboer, 2010). This helps guide instructors in how to “adjust their instruction in accordance with the progress of the class” (Ashenafi, 2017, p. 226).

2.4.2 Effectiveness of PA. Researchers from all over the world have conducted rigorous analyses on the effectiveness of PA. There are three prominent meta-analyses on the reliability of PA. Topping (1998) reviewed 109 research papers on PA in higher education. He reported 11 main parameters that might influence the effectiveness of PA reflected by the previous literature. He argued that both “simple quantitative feedback”, namely numerical scores and grades, and “detailed feedback” have “positive formative effects” about learning gains and learners' attitudes (p. 267). This argument sheds light on the robust formative potential PA entails even with limited feedback available. Moreover, he also reported that a significant number of studies on PA in writing classrooms indicated that peer feedback holds comparable reference value as teacher feedback and sometime it is even more informative than teacher feedback. Different from the qualitative method and descriptive approach adopted by Topping (1998), Falchikov and Goldfinch (2000) conducted a quantitative analysis of 48 research papers on PA. The objective of their analysis was to compare the convergence between PA and TA. By referring to seven major parameters influencing PA, they concluded that PA and TA results are equivalent concerning overall evaluation by explicitly-explained rubrics and intellectual works.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

Most recently, Li, Xiong, Zang, Kornhaber, Lyu, Chung and Suen (2016) identified that neither Topping (1998) nor Falchikov and Goldfinch (2000) included enough research papers on computer-assisted PA, although the rise of the digital age prompted a remarkable quantity of research in this realm. They found a global Pearson correlation of 0.63 between PA and TA as opposed to 0.69 as identified by Falchikov and Goldfinch (2000). The decrease of 0.06 correlation coefficient indicates that there are new factors influencing the effectiveness of PA. Among the nine factors they identified, it is particularly interesting to note that PA has a closer correlation to TA when it is paper-based instead of computer-based and when participants are randomly assigned to groups. Also, PA is more effective when scores and comments are both accessible to learners. The reason for the lack of effectiveness in computer-assisted PA is believed to be that the use of computers causes reduced attention among students and limited spontaneous instruction from teachers (Suen, 2014).

Besides the meta-analyses, there are recent research papers from a great variety of contexts looking into various PA models implemented in higher education. Concerning the convergence between PA and TA, Walker (2015) examined the features and quality of PA on students' written work. The participants were 73 mature students (25 writers and 48 peer reviewers) enrolled in a course with an academic level equivalent to the second or third year of an undergraduate curriculum. The feedback from students on the initial draft and the instructor on the final submission was coded into four categories. Then a comparison within each category between PA and TA was conducted to shed light on the advantages and disadvantages of PA. The results indicate that the feedback offered through PA is different from that through TA but of generally good quality. This is a unique strength of PA as it complements what is missing in TA due to teachers' large workload and time limitation. However, it was found that during the

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

revision stage, students sometimes ignored the feedback from their peers. The mediator role of teachers is therefore significant since teachers can raise students' awareness on valuing peer feedback through spontaneous instruction during the assessment process. This point has also been stressed by Zhao (2014). She realized the significance of teacher mediation in the implementation of PA and proposed an action model to improve the efficiency of classroom PA with specific teacher intervention strategies. 18 English major students participated in PA for nine writing sessions. Using four intervention strategies: assessment training, continuous review of peer feedback, teacher validation, and in-class meditation. Students' PA competence was significantly improved as the divergence between peer feedback and teacher feedback narrowed. The post-assessment surveys demonstrate an overall satisfaction with this action model. The proactive involvement of the teacher in the PA process improved the efficiency of PA and students' judgments on the quality of peer feedback. Interviews with the participants also revealed a positive attitude towards the assessment model. Students expressed that they "benefited from reading peers' writing while responding" and "learned from the shortcomings of peers' writing" (pp. 162-163).

Liu and Li (2014) examined the influence of pre-assessment training on students' assessment competence, and the relationship between students' assessment ability and the quality of peer feedback among undergraduate students in the United States. They received a training session before PA which included clarifying key concepts and rubrics as well as a practicing model. Then they conducted peer assessment on each other's written work for a child education class. The results indicate that training contributes to a much smaller difference between learners' scoring and teacher's scoring.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

With a particular focus on the feedback from PA, Tsai and Chuang (2013) conducted a quasi-experimental study examining the influence of PA on students' argumentative written work. 48 undergraduate students enrolled in two EFL composition classes participated in the study. They were randomly assigned to two groups. The experimental group received a PA training session integrated into the curriculum, and the control group was not exposed to pre-training. At the end of the training session, students from both groups conducted revisions on their written work. The results indicate that the trained group was capable of producing higher frequency and more diverse types of feedback, which led to them producing argumentative writing with better overall quality. The process also helped students to become more critical of their written work and to “invest more effort in spontaneous revision to produce high-quality argumentative writing” (p. 210). They developed the ability to distinguish constructive or non-constructive feedback from their peers.

Concerning the L2 learning context specifically, an important factor to consider is individual difference. It seems noteworthy to look at the role of students with low proficiency and how they contribute to the PA. Yu and Lee (2016) focused on this issue and conducted research in an EFL writing classroom. Forty-one first-year undergraduate students with greatly diversified levels of English proficiency participated in their research. They conducted PA on five short written assignments (220 to 350 words each) in groups of three or four. The results indicate that EFL learners of low proficiency are capable of providing peer feedback on a broad range of peers' written work and of assisting fellow students to improve their writing competence. Several “contextual factors” were identified by the authors to facilitate the positive involvement of low proficiency students in PA, which include “the medium of group discussion, grouping principle and student relationship, low proficiency students' motivation and

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

engagement, and peer feedback training” (p. 491). This study demonstrated that with proper adjustments, the formative function of PA benefits all students.

On the other hand, many researchers measured the effectiveness of PA as reflected by the attitudes of instructors and learners. In America, Carnell (2016) investigated the effectiveness of PA in the third and final undergraduate dissertation course in a UK architectural school. Questionnaires with reflective open-ended questions were used to foster peer feedback. The feedback contains valuable constructive criticism from peer reviewers and revealed that PA had enhanced students’ critical thinking competence which holds great value for students’ academic and professional advancement. Corresponding to Topping (2009), they found that the reciprocal peer review process had benefits for both the assessors and the assessees. As students cultivating “responsible thinking and reflection on a deeper level”, they became more “reflexive and critical of their own work” (p. 1271). Students also expressed a very positive perception towards the assessment process during the post interview.

Ubaque Casallas and Pinilla Castellanos (2016) investigated the influence of PA on the development of students’ argumentation skills. The research project was undertaken by 2 ESL instructors and 12 Colombian students in an intermediate English class. They used audio-taped conversations and open-ended interviews to collect data. The results indicate that PA “promotes learners’ awareness and ability to engage in argumentation process” (p.111). This might be because students were involved in the assessment process, which motivated them to understand the rubrics on a more detailed level as both assessors and assessees. The researchers also claimed that PA is “an essential tool for enhancing personal and collaborative learning, as well as for promoting learner reflection and agency” (p. 119). This statement is consistent with Ashenafi’s (2017) claim that PA is not only an assessment tool but also a learning engine.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

In Asia, Fazel (2015) looked into students' perceptions of PA; specifically, the perceived benefits and weaknesses of PA. The author engaged 42 undergraduate Iranian students majoring in English Translation. The findings from the post-intervention survey indicate an overall positive attitude towards PA as 90.4% students expressed interest in continuing to use it in the future. Students witnessed improved academic presentation skills after the assessment session as well as a general feeling that the course became more interesting than before. Interestingly, six participants remarked that "they felt more empowered by being part of the assessment" (p.86). This response associates PA with learners' self-efficacy.

Jung (2016) compared the discrepancy between PA and TA in the tertiary level EFL writing classroom. The results indicate that scores of TA were lower than that of PA on the first draft while on the second draft the opposite pattern was observed. Overall participants demonstrated a positive attitude towards PA as 69% of them thought that conducting PA was somehow helpful in developing L2 writing.

However, not all research studies demonstrate positive attitudes concerning learners' attitudes towards PA. For example, Vasu, Ling, and Nimehchisalem (2016) investigated ESL students' attitudes towards TA and PA in guiding the revision of written work. Questionnaires were distributed to 107 EFL Malaysian students at the tertiary level. Student responses indicate that TA was perceived as more useful compared to PA in improving their EFL writing skills. The authors analyzed the reason behind this tendency and concluded that students in Malaysia are used to the TA pattern and believe that TA is more reliable in accuracy and more clear in directions for revisions. Students also expressed a preference for concrete and precise explicit feedback, rather than implicit feedback.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

These recent studies are consistent with the previous sections of this literature review in the sense that mixed results are observed concerning the reliability of PA while students' attitudes are generally positive. This is because of the significant variances of PA models implemented in each study and the diverse sociocultural contexts. However, the formative function of PA is never in doubt. All studies agree upon the fact that PA serves as more than an assessment tool in the classroom. More importantly, it should be viewed as a learning tool as well, which is also the perspective of the present study.

2.5 Feedback in FA

Feedback, as defined by Winne and Butler (1994), is “information with which a learner can confirm, add to, and overwrite, tune, or restructure information in memory, whether that information is domain knowledge, meta-cognitive knowledge, beliefs about self and tasks, or cognitive tactics and strategies” (p. 5740). As the learning component of FA (Liu & Carless, 2006), it carries the purpose of deeply involving students in “meta-cognitive strategies such as personal goal-planning, monitoring, and reflection” (Clark, 2012, p. 210). Afitska (2014) further clarifies that feedback generated during FA

... may serve as a supportive bridge which allows learners to move from where they are at the particular moment of their learning to where they are expected to be by their teacher or program. Feedback given as part of formative assessment may help learners become aware of any gaps that exist between their desired goal and their current knowledge, understanding, or skill and guide them through actions necessary to obtain the goal. (p. 30).

The formative nature of feedback is not inherent. It only manifests when feedback represents “information communicated to the learner that is intended to modify his or her

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

thinking or behavior for the purpose of improving learning” (Shute, 2008, p. 154). In other words, the uptake of feedback by learners in the form of immediate responding or delayed revision is what makes it formative (Rea-Dickins, 2003).

According to the Assessment Reform Group (1999), feedback from FA takes two forms: the first form is evaluation-associated remarks, which identify the gap between the learning objective and the current learning status; the second form is feed forward remarks, which point out the required knowledge skill set for the next step of learning process.

However, as Havnes, Smith, Dysthe and Ludvigsen (2012) claimed, “positive effects of feedback are not always the case” (p. 21). Through a two-year intervention project, the authors identified three influential factors in the perception of FA: first, teachers value feedback from FA more than students; second, feedback is generally perceived as more useful in science-focused fields of study; third, female students expressed higher demand about the quality of feedback than their male counterparts. In addition, Cho and Park (2014) also noted that explicit feedback is more effective in cultivating student writers’ awareness of accuracy and clarity than implicit feedback. Moreover, students’ reaction to feedback varies greatly across the literature. Some researchers reported positive uptake of feedback after FA sessions, while others reported that feedback was ignored (Wingate, 2010; Walker, 2015).

Researchers in the field of SA and PA have been advocating for the formative use of feedback from these two assessment approaches. Brown and Harris (2013b) argued that SA should no longer be treated as an assessment, but instead as an essential competence of self-regulation. Topping (2003) pointed out that peer feedback is one of the theoretical underpinnings of PA. Feedback can reduce errors and have positive effects on learning when it is received thoughtfully and positively. The significant deviations between FA and TA revealed by Suñol et

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

al. (2015) demonstrated that formative values of SA and PA should be placed over the summative usefulness of these two evaluation practices.

To date, L2 assessment studies still primarily focus on standardized benchmarks and the role of instructors (Stoynoff, 2012; Turner, 2012), even though feedback and the evaluation process are beneficial for learners as well (Cho & Cho, 2011; Nicol, Thomson & Breslin, 2014). Turner and Purpura (2016) pointed out that L2 assessment studies emphasize grades over learning. The same claim has also been made by other researchers (Brown, Andrade & Chen 2015; Carless, 2011; Way, Dolan & Nicolas 2010). They pointed out that research has acknowledged the central role of assessment in teaching and learning as well as the value of FA and significance of quality of feedback. Nevertheless, the empirical evidence demonstrating the relationship between learning, FA and feedback is still scarce. There is a necessity for more research of FA at the tertiary level connecting the theoretical foundation to empirical evidence (Lau, 2016).

2.6 The Present Study

FA serves dual functions in L2 learning at the tertiary level. As an assessment tool, it provides teachers with a more comprehensive evaluation of learner's academic performance and learning progress, so that timely adjustment can be made to the curriculum. As a learning tool or a pedagogical approach, FA grants students the opportunity to improve their self-regulation and autonomy in the learning process, which will transfer to valuable skills in their future academic and/or professional development.

Two common assessment approaches commonly used for formative purposes, SA and PA, have drawn particular interest from scholars as they have been widely applied to tertiary-level classrooms in various contexts. Although mixed results have been shown through the

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

literature review regarding the reliability of SA and PA as summative evaluation methods, the robust formative function and students' overall positive attitude towards SA and PA implies that they have significant pedagogical potential. However, the literature reveals that studies on SA and PA, to a great extent, focus on comparing the convergence of scoring between teachers' ratings and students' ratings, and there is a lack of attention to the quality of feedback produced through the FA process.

The present study will investigate the effectiveness of FA with the focus shifted from assigned grades to the quantity and quality of feedback. The effectiveness of FA is represented by the errors identified in students' written work as a form of feedback instead of numerical scores. The research study is situated within the CBA research agenda proposed by Turner and Purpura (2016) and furthers the theoretical foundation of how FA contributes to learners' independence and autonomy. Since a formative view is adopted to define SA and PA used in this study, and the assessment practices are integrated into instructed language learning, these practices will be categorized as informal FA. The rationale of the study answers the call for more research studies on the formative purpose of evaluation (Brown, Andrade & Chen 2015; Carless, 2011; Stoyhoff, 2012, Turner, 2012; Way, Dolan & Nicolas 2010) as well as positioning feedback quality in the central of conceptualizing formative assessment practices (Turner & Purpura, 2016). The present study also investigates the characteristics of feedback on students' written work, which will disclose the nature of different FA approaches and give useful implications to teachers' curriculum planning.

2.7 Summary

In this chapter, the researcher first acknowledged the complexity of CBA and reviewed a tentative research agenda of CBA, then situated the present study within the theoretical

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

framework of CBA. Key definitions related to the present study (CBA, LOA, FA and AFL) were clarified with explanations of their relationships. Following that, a detailed review of the two assessment approaches used for formative purposes in the present study was presented with empirical evidence from a large variety of contexts. This led to a discussion of the key parameter used in this study which was the feedback produced through FA. The significance and central role of feedback were analyzed and supported with previous literature. By doing so, a need for shifting the focus of research from examining numerical scores to understand the nature of constructive feedback was elicited.

In the next chapter, the methodological rationale of the present research will be presented in several categories, including the research question, research context, research design, data collection methods, and data analysis procedures.

Chapter 3: Methodology

3.1 Introduction

This chapter addresses the methodological underpinnings of the present study. It begins with the elicitation of two primary research questions. Following that, the research setting and participants are briefly described. Ethical considerations in accordance with the regulations of McGill University's Research Ethics Board (REB) are presented. Following that, the three-session research design and data collection instruments are discussed in detail. Then a review of quantitative data analysis procedures applied in this study is presented followed by a short summary at the end.

3.2 Research Questions

The purpose of this quantitative study is to investigate the quantity and type of feedback provided through formative SA and PA as pedagogical practices in English L2 writing classrooms.

Accordingly, the research questions are:

1. How much feedback do students give on written assignments during SA, PA and a combination of both SA and PA (henceforth referred as CA)? Specifically, do students produce more feedback through CA than through SA or PA alone? (RQ1)
2. What aspects of written assignments (content, organization, vocabulary and grammar) do students give feedback on within these different types of assessment? Specifically, which aspects of written assignments do students give more feedback on within SA? Which aspects of written assignments do students give more feedback on within PA? (RQ2)

To address these two questions, the research process involves an interactive workshop, a composition writing and a FA session. However, the quantitative research data is only collected

during the FA session.

3.3 Research Setting

Jining Medical University is a higher education institution and medical center in southwestern Shandong Province, China. It has approximately 13500 undergraduate students. The present study took place at the School of Foreign Languages and all student participants were freshmen and sophomores major in English Language and Literature.

The two courses in which this study took place were an Integrated English Course (for first year undergraduate English majors) and an English Writing Course (for second year undergraduate English majors). These classes were selected because the designated L2 writing class is not available for first year English majors; however, both classes had an academic English writing teaching component. Consistent with the general tertiary-level L2 instruction perception, their syllabuses demonstrate a language-driven, form-focused and test-oriented learning environment. The primary teaching objective of the Integrated English Course is to enrich students' vocabulary and grammar knowledge in preparation for more advanced classes; while the English Writing Course is specifically designed to prepare students for a national standardized high-stakes test: Test for English Major (TEM) Level 4.

For English majors in Chinese colleges and universities, TEM is an important test because passing it is a compulsory requirement for obtaining a Bachelor of Arts degree. The test, which is administered by the National Advisory Commission on Foreign Language Teaching in Higher Education (NACFLT) on behalf of the Higher Education Department, Chinese Ministry of Education, is held annually in every higher education institution that offers an English major. There are two levels in TEM, level 4 is conducted at the end of the second year, and level 8 at the end of the fourth year in their undergraduate program (Cheng, 2008).

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

The purpose of the TEM is to measure the English proficiency of Chinese university undergraduates majoring in English Language and Literature and to examine whether these students meet the required levels of English language abilities as specified in the National College English Teaching Syllabus for English Majors (NACFLT, 2007). There is also a formative side of the test since the assessment results are used for improving English instruction and acquisition at the national level. TEM is a criterion-referenced test, which means “students’ performance is evaluated against the criteria stipulated by the teaching syllabus” (Cheng, 2008, p. 20). The reliability and validity of the test system has been widely recognized by test takers, educational institutions and the general public and is closely related to English major students’ employability in the workplace. At the time of this study in Jining Medical University, participants were at the beginning stage of preparing for the TEM Level 4 (TEM 4).

Since its debut in 1992, TEM 4 has undergone several major revisions. The most recent one took place in 2016, where the format of all four parts of the test have been changed. The content and format of TEM 4 after the revision is shown in the following Table 1:

Table 1

Test content and format of TEM 4 adapted from Jin and Fan (2011)

Task	Input	Format	%	Length (min)
Listening	1 passage, listen four times, 80 words, 120 wpm	Word-for-word dictation	10	10
	2 lectures, listen once, 500 words, 120 wpm	Filling blanks	10	10

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

	2 conversations, listen once, 250 words each, 120 wpm	Multiple choice	10	10
Reading	4 to 5 texts totaling 1800 words	Multiple choice + Short answer	20	35
Grammar & Vocabulary	20 sentences	Multiple choice	20	10
Cloze	One text, 250 words	Multiple choice	10	10
Writing	Written prompts	A text of 200 words	20	45
Total			100	130

The reformed written prompt entails two synthesized tasks, a summarization of information from an excerpt followed by an argumentative essay on a contemporary sociocultural topic related to the excerpt. It lasts for 45 minutes and the total word count requirement is 200 words. Holistically, students' writing competence is assessed against the criteria of content (relevance and completeness) and language (grammar accuracy and vocabulary appropriateness) (NACFLT, 2015).

As a section that occupies the longest time slot in the TEM4 and constitutes 20% of the total marks, the writing section is considered challenging by many test-takers (Chen et al., 2014). As pointed out by Jiang (2013), the reasons for Chinese students' low achievement in academic writing are three-fold. First, Chinese and English have different rhetorical strategies. Chinese students tend to apply rhetorical conventions taken directly from the linguistic norms in Chinese, which results in digression from the topic. Second, Chinese students tend to use excessive

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

proverbs in academic English writing. Although commonly used in Chinese tests to support an argument, it is not a proper practice to insert proverbs and sayings in English academic writing. Thirdly, Chinese students show less awareness on language cohesion. There is a lack of sufficient organizing devices in Chinese students' writing, for example transitional adverbs and relative clauses which contribute to a higher level of clarity. However, with sufficient and proper preparation guided by instructors, a large proportion of English major students are capable of passing the test, which means achieving at least 60% on the test. The percentages of students passing the TEM4 in recent years are demonstrated in Figure 3. Although the TEM 4 passing rate in Jining Medical University fluctuates over the years, it is always higher than the national average level.

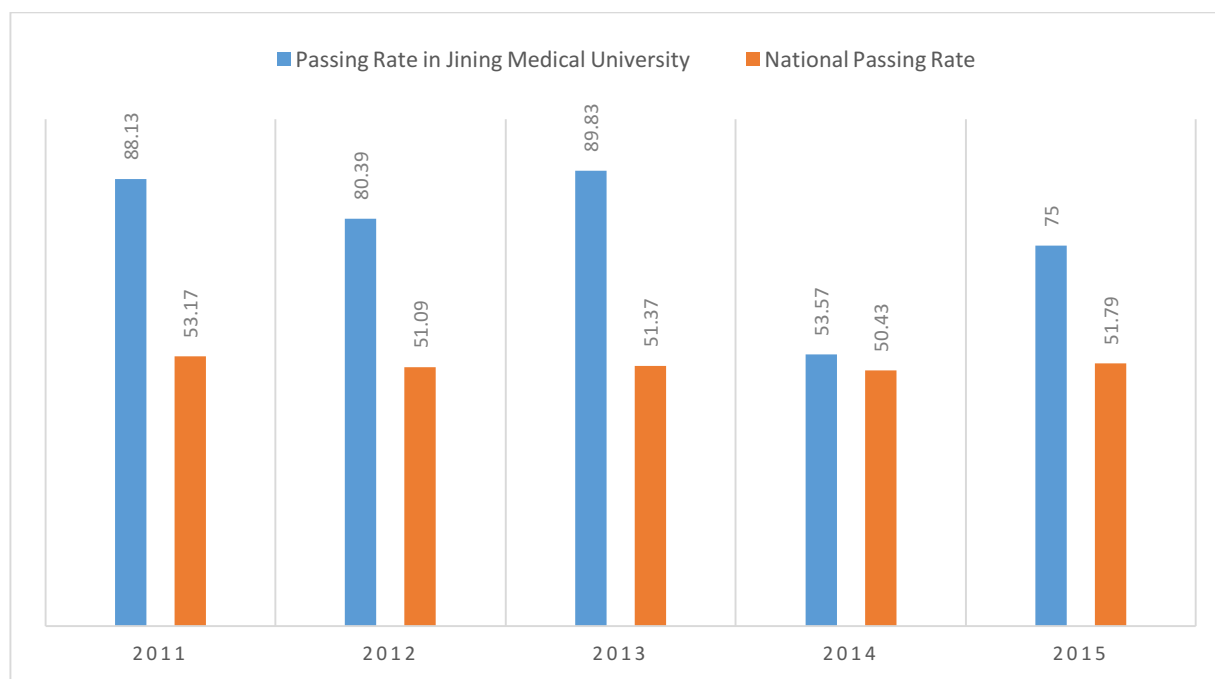


Figure 3. *TEM4 passing rates comparison (2011-2016)*

3.4 Participants

Participants in the study could be categorized with three labels: student, teacher and reviewer. There were 97 students in total who participated in the present study. The proportion of

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

female students was significantly larger than that of their male counterparts. The age range of students is 18 to 23. All students were first or second year undergraduates from Jining Medical University majoring in English. They enrolled in the university in 2016 and 2015. Students enrolled in each year were randomly divided into two classes after registration. Therefore, four classes are involved in the present study. The detailed demographic information is presented in the following Table 2.

All students were Mandarin L1, and they had been learning English as a foreign language with low time intensity (two 40-minute class sessions per week during academic years) from Grade 3 to Grade 12. All participants were engaged in the three research sessions; however, due to the lack of signed consent forms or valid data (i.e. blank reflective sheet, quantity of feedback ≤ 3), 79 pieces of data were collected and analyzed.

Table 2

Demographic information of student participants

Admitted Year and Class #	Male	Female	Total
2015(1)	4	16	20
2015(2)	3	20	23
2016(1)	4	22	26
2016(2)	4	23	27
Total	16	80	97

Participating students were selected for admission through the National College Entrance Examination, which ensures that their general English proficiency level was intermediate. They

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

had had very limited exposure to formative assessment in English writing classrooms and TEM 4 since they were still at the beginning stages of test preparation when this study took place; however, they are aware of the close relationship between TEM 4 and their academic progress as it has been discussed during university orientation.

The two teachers (T1 and T2) involved in the research were responsible for the instruction of the Integrated English Course and the English Writing Course respectively. Both had been teaching at Jining Medical University for over five years. Their pedagogical practices are predominantly test-oriented and form-focused, and do not involve FA practices. The reviewer in this study is also an instructor at the same university but has no classroom contact with the student participants. She holds a Master in Second Language Education from University of York in Great Britain and has been an in-service instructor for 10 years. There is no perceived conflict of interest between the reviewer and the students.

3.5 Ethical Considerations

In compliance with the McGill Research Ethics Board (REB) rules, regulations, and processes regarding research with human subjects, the researcher obtained access to the research site and consent from the participants. A research proposal, containing informed consent forms and research designs, was presented and approved through the review of REB prior to the data collection stage.

Specifically, in summer 2016, initial contact to the administration department of Jining Medical University was made via email (See Appendix A: Initial Communication). The researcher briefly introduced the intended research project and how all participants would be involved in the research procedures. The theoretical contributions and empirical benefits of the study were presented as persuasive arguments for promoting participation. Then a request of

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

permission to enter the classrooms and to implement the intervention protocols was proposed. In late September, 2016, the university granted the researcher the access to the target classes and potential participants.

There are several course sessions in both Integrated English and English Writing Curriculum with no specific teaching objectives. They only serve as additional independent study sessions for students. The research study took place during these class sessions, so students' regularly scheduled class time was not disturbed. The interactive workshop implemented in the research was non-optional, and was integrated as part of one self-study class session which all students were required to attend. The workshop was intended to further ensure that participants are equipped with proper skills to practice formative assessment in the subsequent research sessions. Therefore, no research data was collected at this stage. After the workshop, students decided whether they wanted to participate in the subsequent sessions. Those who decided to participate signed a consent form (See Appendix B: Student Consent Form), while those who decided to opt out were still able to participate in and benefit from all the research activities (essay writing, SA and PA), but their data were not included in the study. Besides being informed about the following research procedures, participants were also assured that they reserved the right to withdraw from the study and any time with no consequence. As for teacher participants, given the fact that only a limited number of teachers were available for the research study, they were contacted directly via in-person meetings. Participating teachers were asked to sign a separate consent form (See Appendix C: Teacher Consent Form).

3.6 Research Design

This study adopts a three-group posttest-only randomized experimental design. The research process consists of three sessions conducted over two weeks.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

In the first session, the non-optional 60-minute interactive workshop was given as part of one course session to potential student and teacher participants on October 17th, 2016. The workshop was coordinated by the researcher and presented the core concepts of FA, the rationale of the present study, and how to engage in SA and PA. The purpose of this workshop was to motivate participation and to ease potential participants' concerns about the feasibility and ethics issues in the research. At the end of the workshop, students were guided to review a practical model of FA, which served as a training session for students. This process was to ensure that participants were equipped with proper skills to practice FA in the following phases. Therefore, no research data was collected at this stage. After the workshop, students and teachers decided whether they want to participate in the research study or not.

The second session took place on October 18 and 19, 2016. It consisted of three parts conducted over 75 minutes. The first part consisted of a 20-minute overview of the analytic rubrics (See Appendix D: Analytic Rubrics), which was developed by the researcher based on the holistic evaluation criteria of the writing section in TEM 4 and a five-step guide on developing rating rubrics proposed by Crusan (2010).

The analytic rubrics had four sections: content, organization, grammar and vocabulary. Each section contained specified rules of academic English writing and was explained to students with examples to ensure accurate interpretations could be made during the formative assessment process. In other words, students were instructed to identify issues with accuracy of language forms as well as other areas of improvement based on the analytic rubrics.

Although numerical scales were used in the development of the analytic rubrics, in an effort to encourage students to give descriptive feedback, grading and ticking boxes were deliberately avoided in the present study.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

Next a writing prompt was distributed to students (See Appendix E: TEM4 Writing Prompt) and a guided task reading was conducted by the researcher. The writing prompt was adapted from 2016 TEM 4 test paper, which contains a general description of the writing task (i.e. format, length and time) and an excerpt with relevant background information. The readability of this excerpt has been adjusted to Flesh Kincaid Grade 8, which is comprehensible to 80% of Americans; therefore, the influence of students' English proficiency levels on task comprehension was hedged. The guided task reading, during which the researcher read and explained the writing task to students, is designed to ensure that students fully understand the two synthesized tasks which is a summarization of information from the excerpt and an argumentative response to the central topic in the excerpt since they had not practiced these types of writing tasks prior to this research study.

Lastly, a 45-minute independent writing was implemented and all students were required to write a response to the writing prompt with no less than 200 words. In summary, the three-step process of the second session is demonstrated with the following Figure 4.

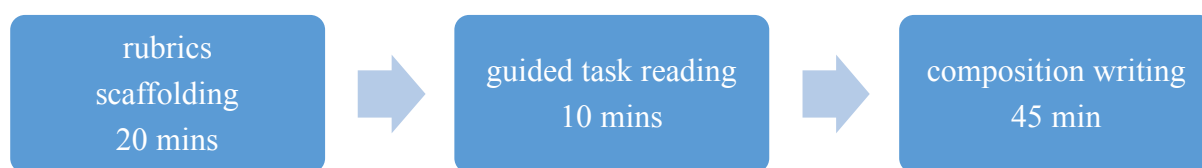


Figure 4. *Three-step process of the second research session*

The third session started with in-class SA, PA and CA conducted by students. This process lasted for 45 minutes. After the composition writing, students were randomly assigned into three groups. Each group conducted SA, PA and CA respectively. The specific group plan is presented in Table 3.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

Table 3

Student participants group plan

	SA Participants	PA Group #/ Participants	CA Group #/ Participants
2016 (1)	10	Group1/ 3 Group2/ 3 Group3/ 4	Group1/ 3 Group2/ 3 Group3/ 3
2016 (2)	9	Group4/ 3 Group5/ 3 Group6/ 3	Group4/ 3 Group5/ 3 Group6/ 4
2015 (2)	6	Group7/ 3 Group8/ 3	Group7/ 3 Group8/ 3
2015 (1)	4	Group9/ 3 Group10/ 3 Group11/ 3 Group12/ 3	Group9/ 3 Group10/ 3 Group11/ 3 Group12/ 3
Total	29	34	34

As shown in Table 3, a total number of 29 students were assigned to the SA group and instructed to conduct criterion-referenced evaluation on their own compositions. 34 students were assigned to 12 groups of three or four conducting PA. Each student in one group was asked to evaluate two peers' drafts. Similarly, 34 students were assigned to 12 groups of three or four conducting CA. Each student in one group was required to evaluate their own written responses and then evaluate one peer's work. Spontaneous instructions were provided to all students based on voluntary verbal requirements. For instance, when students in one group disagree with each

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

other on specific grammatical rules, the research provided necessary verbal explanations to facilitate the assessment process.

Following that, reflective sheets (See Appendix F: Reflective Sheet) were distributed to students. On these sheets, they were instructed to write down the various types of errors in their compositions identified by either themselves or their peers under four categories corresponding to the analytic rubrics, namely content, organization, grammar, and vocabulary. Students were strongly encouraged to document the identified errors in a descriptive format pointing out the line in which the error was located and how the error could be corrected. This process is crucial to the methodological soundness of the present study as it represents the shift from assigning numerical scores to generating and analyzing feedback. The errors identified by students through FA were marked as student-identified errors (SIE).

These reflective sheets were then collected and handed to the teachers. T1 and T2 reviewed the SIE and determined whether the feedback given by students aligned with lexical, syntactic, logic and stylistic rules in academic writing. The SIE confirmed as correctly identified errors by teachers were coded as SCIE. SCIE were considered to be formative feedback in this study as they had the potential to improve students' written work if used in the revision process. Finally, all students' compositions and the reflective sheets were collected and handed to the reviewer who conducted TA according to the same analytic rubrics presented to students. Then all errors identified by the reviewer in students' written work were transcribed to the reflective sheets under the four categories. These errors identified by the review were marked as overall errors (OE) in data coding. The frequency of OE on all students' written work were treated as the baseline in this study for data manipulation and cross-group comparison. The reviewer finished evaluating the compositions on Oct. 30th, which ended the third session of the study.

In summary, the process of data collecting and coding is demonstrated in the following Figure 5.



Figure 5. *Process of data collecting and coding*

3.7 Data Collection Instruments

The quantitative data in the present study come from the FA practices used in the third session. The procedures of SA and PA applied in the present study are developed on the basis of four references. Topping (2009) offered an eleven-point guidance on organizing PA in classrooms, based on the context of the present study, the following six pieces of guidance were adopted based on the research context:

1. Clarify purpose, rationale, expectations, and acceptability with all stakeholders.
2. Provide training, examples, and practice.
3. Provide guidelines, checklists, or other tangible scaffolding.
4. Specify activities and timescale.
5. Monitor and coach.
6. Examine the quality of peer feedback. (Topping, 2009, pp. 191-192)

Crusan (2010) proposed a review sheet for PA which consists of 10 open-ended questions prompting students' descriptive response. These questions are used in facilitating the interactive workshop, eliciting the analytic rubrics and spontaneous instruction during the FA session.

As for SA, Nielson (2014) summarized twelve strategies from previous research on implementing effective SA in writing classrooms. Given the context of the present study, the following 5 strategies were adopted based on the research context:

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

1. Provide direct, explicit and step-by-step training in the use of self-assessment exercise beforehand; include practice time prior to implementation.
2. Teach students the criteria for rating their own work and the evaluation of specific aspects of writing.
3. Students should be provided sufficient independent time and space to complete self-assessment exercises, with most work taking place in class.
4. Students need teacher support, monitoring and feedback during the self-assessment exercises. Support should include student-teacher dialoguing.
5. Self-assessment should include tasks that address specific components of writing, as well as holistic prompts eliciting global response. (Nielson, 2014, pp. 9-11)

Since SA and PA are closely related evaluation approaches with shared formative function (Carnell, 2016). Similar designs for SA and PA are adopted in this research. The procedures are demonstrated in the Figure 6 below:

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

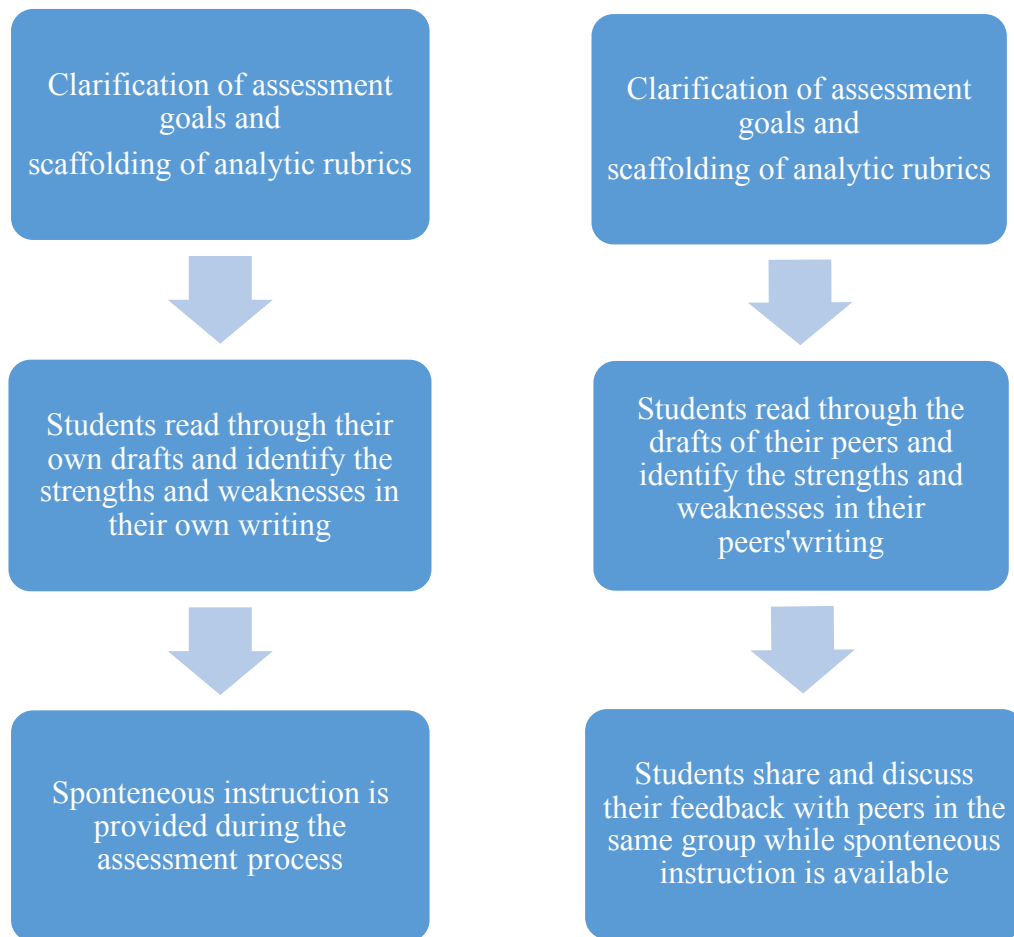


Figure 6. *Procedures of SA (left) and PA (right)*

Anderson (2012) proposed an action model for CA, a group work evaluation form consisting of six questions is used to guide the assessment activity. The first three questions and the last question are used to prompt peer feedback while the other two questions are used to foster self-reflection. This action model is adopted to guide the CA in the present study. When students are assigned to conduct CA, it means that they are asked to read the drafts of their own and one draft written by one of the peers in the same group; they are encouraged to discuss about the strengths and weaknesses they identified in either their own compositions or their peers' compositions in the format of a group discussion. Therefore, students have access to more opportunities to reflect on all the written work they were exposed to.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

After the FA, students are then asked to transfer the errors they identified to a reflective sheet in a descriptive manner. The frequency of feedback given in the reflective sheets is the main source of quantitative data collected in the present study.

3.8 Data Analysis

The quantitative data collected in this research were analyzed with SPSS on two levels in order to answer the proposed research questions: descriptive statistics and cross-group statistical analysis. On the descriptive level, frequency tables were used to present the synthesized raw data and generated descriptive conclusions based on observation. On the statistical level, both Analysis of Variances (ANOVA) and Mann-Whitney U test were used based on the distribution characteristics of the data sets. With the two-level data analysis, a well-supported and comprehensive understanding of the research data was established.

3.9 Summary

In this chapter, the research questions corresponding to the scholarly gaps identified in the literature review section were presented. Then, research background was discussed, including the participating university, the participating classes, and the high-stakes large-scale test related to the research design. Following that, ethical considerations were presented as human participants were involved in the research process. Finally, the three-session research design, the quantitative data collection instruments and two-level data analysis process were discussed in detail. In the next chapter, the research results from each level of data analysis will be presented and discussed.

Chapter 4: Results

4.1 Introduction

In this chapter, the quantitative data collected through the FA session will be presented and analyzed. The quantitative data in the present study were transferred into SPSS where all the statistical analyses were conducted. Results will be organized by the relevance to the two research questions proposed in the methodology chapter. Specifically, for research question one,

RQ 1: How much feedback do students give on written assignments during SA, PA and CA?

-Specifically, do students produce more feedback through CA than through SA or PA alone?

Three descriptive cross tables were used to generate observational conclusions about the differences in the effectiveness of different forms of FA; then a one-way ANOVA test was conducted to further examine the statistical significance of the observed differences. For research question two:

RQ 2: What aspects of written assignments (content, organization, vocabulary and grammar) do students give feedback on within SA and PA?

-Specifically, which aspects of written assignments do students give more feedback on within SA?

-Which aspects of written assignments do students give more feedback on within PA?

Three descriptive cross tables were used to generate observational conclusions about the differences in the focuses of feedback, namely content, organization, vocabulary and grammar, in PA and SA; then a Mann-Whitney U test was conducted to investigate the statistical

significance of the observed differences. The connections between the results and the research questions are briefly discussed, and a summary concludes this chapter.

4.2 RQ1: Descriptive Results

RQ 1: How much feedback do students give on written assignments during SA, PA and CA? Specifically, do students produce more feedback through CA than through SA or PA alone?

The quantitative data are categorized into three groups. The first is the number of errors identified by students (SIE) through SA, PA or CA. Secondly, teachers examined the errors identified by students and confirmed whether they were, indeed errors. These students-correctly identified-errors (SCIE) represent the second group of data. Third, the numbers of overall errors identified by the reviewer (OE) were collected as the independent variable for cross-group comparison.

The effectiveness of each type of FA in this study is represented by the discrepancy between SCIE and OE. The discrepancy is computed by subtracting the former from the latter. Therefore, when the data is negative, it means that student participants failed to correctly identify as many errors as the reviewer. Three sets of descriptive data were generated with SPSS and presented in the following Table 4.

Table 4

Descriptive presentation of SCIE-OE discrepancy

FA	N	Mean	Median	SD	Minimum	Maximum
PA	26	-8.69	-8.00	5.03	-21	-3
SA	27	-9.11	-10.00	5.53	-17	0

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

CA	26	-9.19	-9.50	4.92	-18	-2
----	----	-------	-------	------	-----	----

As indicated by the negative means and medians in Table 4, generally students in the present study are not able to correctly identify all the errors through any form of FA; however, in SA, there is at least one student who correctly identified as many errors as the reviewer did as indicated by the maximum 0. In general, student participants in PA groups performed better at correctly identifying errors with the mean of -8.69 which is slightly higher than -9.11 in SA groups and -9.19 in CA groups. With regard to the consistency, student participants performed most consistent in CA (SD= 4.92) as compared to PA (SD= 5.03) and SA (SD= 5.53). However, it should be noted that differences across groups are not significant.

Table 5 shows the proportions of SCIE out of OE in students' written work broken down by error types. The valid 0% column represents the percentages of students who did not correctly identify any errors in the evaluation of certain error types while the reviewer did; the valid $\geq 50\%$ represents the percentages of students who correctly identified at least half of all errors as identified by the reviewer. These two columns offer valuable reference on the effectiveness of FA as the former indicates that participants failed to perform FA effectively while the latter indicates the opposite.

Table 5

Descriptive presentation of SCIE/OE percentages by error types

Error	FA	N	Mean	SD	Valid	Valid
Types	Type				0%	$\geq 50\%$
Content	PA	26	40.6%	45.7%	30.8%	42.4%

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

	SA	27	42.7%	41.2%	25.9%	37.0%
	CA	26	28.1%	37.7%	46.2%	30.6%
Organization	PA	26	65.3%	62.1%	29.2%	54.1%
	SA	27	67.6%	69.4%	37.0%	59.2%
	CA	26	44.2%	75.5%	56.5%	30.2%
Vocabulary	PA	26	68.4%	54.1%	16.7%	54.2%
	SA	27	73.3%	81.6%	22.2%	51.8%
	CA	26	54.0%	42.7%	20.0%	52.0%
Grammar	PA	26	31.7%	32.0%	26.9%	27.8%
	SA	27	37.8%	36.6%	18.5%	33.3%
	CA	26	38.0%	30.8%	19.2%	34.5%

As shown in Table 5, the means of percentages indicate that for all assessment types, students found fewer grammatical errors than other types of errors. It also shows that CA worked differently for grammar than for other error types. Specifically, students appear to correctly identify the largest proportion of errors through SA on the content, organization and vocabulary with 42.7%, 67.6% and 73.3% respectively. In contrast, students only correctly identified 28.1%, 44.2% and 54.0% of OE on the same error types through CA, which are the lowest percentages among the three forms of FA implemented in the present study. Concerning grammatical errors, it seems that students correctly identified the largest proportion of errors through CA. However, this percentage is significantly lower than the lowest percentages in other three error types.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

Although students appear to correctly identify the most vocabulary errors through SA, it should also be noted that the standard deviation in the same category is also the highest of all the FA variations with 81.6%; and the lowest standard deviation occurs in CA for grammar.

This leads to the last two columns in Table 5. 46.2% and 56.5% of students failed to correctly identify any errors in the evaluation of content and organization through CA, which are the highest among three FA forms; while 59.2% of students correctly identified at least 50% of all the organizational errors as identified by the reviewer through SA and 42.4% of all contextual errors through PA representing the highest proportion on the evaluation of organization and context respectively. In general, fewer students fail to correctly identify any lexical and grammatical errors with the largest proportion being 22.2% through SA and 26.9% through PA. With regard to the evaluation of grammar, the highest proportion of students who correctly identified at least 50% of all errors is only 34.5% which is the lowest compared to the other three error types.

To examine the differences in the effectiveness of the three FA forms more comprehensively and precisely, another way of presenting the differences is shown in Table 3. It adopts the same calculation previously used to generate data sets in Table 1 but breaks the SCIE-OE discrepancy by error type.

Table 6

Descriptive presentation of SCIE-OE discrepancy by error types

Error Types	Assessment Type	N	Mean	Median	SD	Range
Content	PA	26	-2.69	-2.00	2.259	8

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

	SA	27	-2.89	-3.00	2.309	8
	CA	26	-3.04	-3.00	1.843	7
Organization	PA	26	-0.77	-0.50	1.275	4
	SA	27	-1.07	-1.00	1.615	6
	CA	26	-1.23	-1.00	1.632	6
Vocabulary	PA	26	-1.12	-0.50	1.751	7
	SA	27	-1.44	-2.00	1.968	7
	CA	26	-1.27	-1.00	1.589	6
Grammar	PA	26	-4.12	-4.00	2.688	9
	SA	27	-3.70	-4.00	2.799	10
	CA	26	-3.65	-3.00	2.331	9

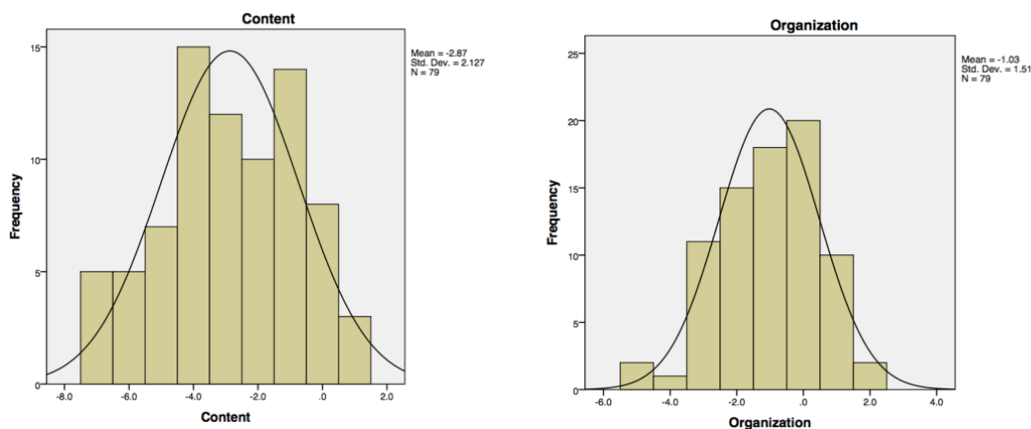
As indicated by the negative means and medians in Table 6, generally students in the present study are not able to correctly identify all errors through any form of FA; however, there is a small difference between different error types of written work. Students in all forms of FA performed better at correctly identifying errors with regard to the organization and vocabulary error types compared to the content and grammar error types. Specifically, PA appears to be the most effective type of FA with the SCIE-OE discrepancy of -2.69, -0.77 and -1.12 on the error types of content, organization and vocabulary; while on the grammar error type, CA seems to be the most effective type with the SCIE-OE discrepancy of -3.65 compared to -4.12 through PA and -3.70 through SA.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

Moreover, students seem to have the most varied accuracy in correctly identifying grammatical errors with standard deviations of 2.688, 2.799 and 2.331 in PA, SA and CA respectively. In contrast, their formative evaluation performance is most consistent across FA type in their evaluations of organization, with standard deviations of 1.275, 1.615 and 1.632 in PA, SA and CA respectively. The statistical analysis and results will be described in the following section.

4.3 RQ1: Statistical Analysis

The testing of the hypothesis used in the present study aims to determine where the differences observed in the descriptive data are significant. In order to conduct a robust one-way ANOVA test, the research data has to meet three assumptions (Phazur & Schmelki, 2013; Turner & Thayer, 2001): first, there must be independent observations. Each student participant in the present study only participated in one and only one form of FA and had no chance to produce data under another form of FA. Second, there must be normally distributed variables, as presented in the following Figure 7, the distribution of all four sets of data complies with a normal distribution.



THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

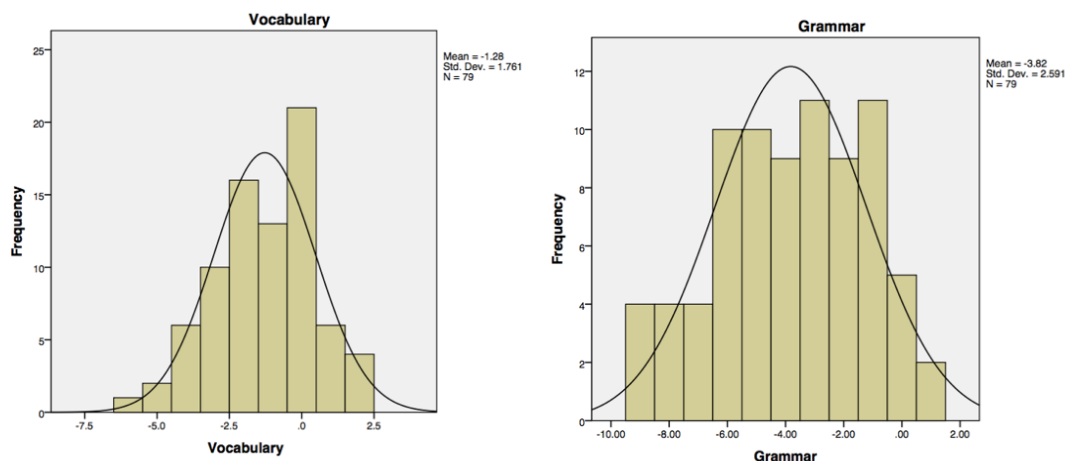


Figure 7. Split histograms of SCIE-OE discrepancy in four error types of written work

Last, there must be homoscedasticity, which means the population variance should be equal across all groups being examined. To determine this, the collected data was examined through the test of homogeneity of variances (also known as Levene's test of equality of error variances), the results are shown in Table 7.

Table 7

Test of homogeneity of variances results

Error Types	Levene Statistic	df1	df2	Significance
Content	.834	2	76	.438
Organization	.512	2	76	.602
Vocabulary	.860	2	76	.427
Grammar	.491	2	76	.614

As shown in Table 7, the significance values (p) for all four error types far exceed the critical value of .05, which indicates the four groups of data satisfy the assumption of ANOVA

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

test which is a relatively equal variance. Therefore, the three assumptions of a robust one-way ANOVA test have all been satisfied.

Table 8

One-way ANOVA test results

Error		Sum of	df	Mean	F	Sig.
Types		Squares		Square		
Content	Between Groups	1.568	2	.784	.170	.844
	Within Groups	351.167	76	4.621		
	Total	352.734	78			
Organization	Between Groups	2.867	2	1.433	.622	.539
	Within Groups	175.083	76	2.304		
	Total	177.949	78			
Vocabulary	Between Groups	1.438	2	.719	.227	.797
	Within Groups	240.436	76	3.164		
	Total	241.873	78			
Grammar	Between Groups	3.351	2	1.675	.245	.783
	Within Groups	520.168	76	6.844		
	Total	523.519	78			

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

Table 8 presents the results from the independent-samples ANOVA test across PA, SA and CA on four error types of written work. There is no statistically significant difference between group means, which means the effectiveness of PA, SA and CA, as represented by SCIE-OE discrepancy, shows no significant different in the present research context. Specifically, in the error type of content, $F(2,76) = .170$, $p = .844$; in the error type of organization, $F(2,76) = .622$, $p = .539$; in the error type of vocabulary, $F(2,76) = .227$, $p = .797$; in the error type of grammar, $F(2,76) = .245$, $p = .783$.

4.4 RQ2: Descriptive Results

RQ 2: What aspects of written assignments (content, organization, vocabulary and grammar) do students give feedback on within SA and PA? Specifically, which aspects of written assignments do students give more feedback on within SA? Which aspects of written assignments do students give more feedback on within PA?

To examine the differences of feedback produced by through PA and SA. First a frequency counts cross table (See Table 9) was built.

Table 9

Descriptive presentation of the frequency counts of SCIE in SA and PA

Error Types	FA Types	N	Mean	Median	SD	Range
Content	PA	26	1.04	1.00	.824	2
	SA	27	1.30	1.00	1.203	5
Organization	PA	26	.92	1.00	.796	3
	SA	27	1.00	1.00	.961	3

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

Vocabulary	PA	26	1.38	1.50	.898	3
	SA	27	1.52	1.00	1.221	5
Grammar	PA	26	1.46	1.00	1.240	4
	SA	27	1.81	1.00	1.520	6

Table 9 presents the descriptive frequency counts of TCE through SA and PA in the evaluation of content, organization, vocabulary and grammar. As indicated by the means and medians, the average numbers of errors correctly identified by students through PA and SA are quite similar, with the most noticeable difference being .35 for grammar. On the evaluation of organization, the difference comes as close as .08 between PA and SA. These differences indicate that in the evaluation of content and grammar, student participants correctly identified more errors than in the evaluation of organization and vocabulary. However, students' competence in correctly identifying grammatical errors varied the most among all the categories with a highest standard deviation of 1.240 in PA and 1.520 in SA.

Table 10

Descriptive presentation of SCIE in each error type of written work/SCIE in total

Error Types	Assessment Types	N	Mean	Median	SD	Range
Content	PA	26	22.6%	20.0%	23.6%	100%
	SA	27	22.5%	22.2%	18.6%	71%
Organization	PA	26	17.5%	18.3%	14.5%	50%

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

	SA	27	16.8%	14.3%	16.0%	50%
Vocabulary	PA	26	29.9%	26.8%	24.4%	100%
	SA	27	29.1%	25.0%	24.4%	100%
Grammar	PA	26	30.0%	25.0%	27.5%	100%
	SA	27	31.2%	33.3%	26.2%	100%

Another way of comparing the different focuses of feedback is the percentages of SCIE in each error type of written work/ SCIE in total, as presented in Table 10. The smallest difference occurs for content with errors identified through PA constituting 22.6% and through SA constituting 22.5% of the SCIE. This contradicts the observations made through the frequency counts in Table 9. However, similar to the observations in Table 9, grammar appears to be the error type with the largest difference as represented by percentages. The difference between the proportions of grammatical errors correctly identified by students through PA and SA is 1.2%.

Moreover, there are significant differences in the ranges of proportions with at least 50% in both PA and SA in the evaluation of all four error types. This indicates that some student participants only identified certain error types, especially on the evaluation of vocabulary and grammar regardless of the type of FA adopted.

Table 11

Descriptive presentation of SCIE/OE total

Error	Assessment	N	Mean	SD	Range
Types	Types				

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

Content	PA	26	9.4%	8.4%	33.3%
	SA	27	10.0%	9.4%	33.3%
Organization	PA	26	7.9%	8.0%	33.3%
	SA	27	8.6%	9.6%	33.3%
Vocabulary	PA	26	11.4%	10.5%	50.0%
	SA	27	12.5%	12.2%	55.6%
Grammar	PA	26	11.7%	11.1%	37.5%
	SA	27	12.4%	10.4%	26.4%

Finally, the ratio between the numbers of SCIE and the total numbers of OE in the evaluation of content, organization, vocabulary and grammar is calculated to represent the varied focuses of feedback produced through PA and SA (Table 11). In general, the percentages are higher in SA than in PA on all four error types of written work. The larger proportion of errors have been identified by students in the evaluation of vocabulary and grammar; specifically, 11.4% and 12.5% of lexical OE have been correctly identified by students through PA and SA respectively; 11.7% and 12.4% of grammatical OE has been correctly identified by students through PA and SA respectively. However, the standard deviations are also higher concerning these two error types ranging from 10.4% to 12.2% as compared to the standard deviations on the other two error types which range from 8.0% to 9.6%.

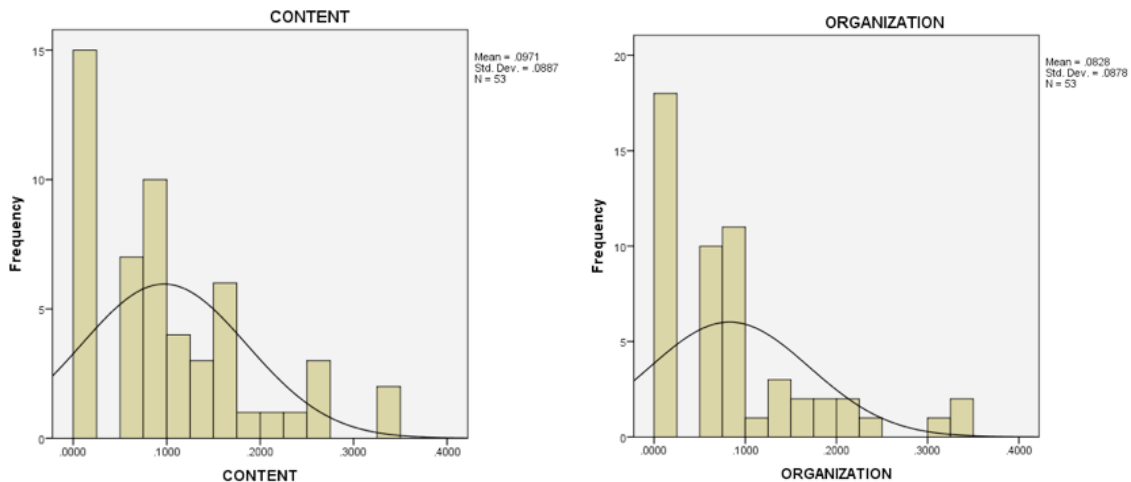
Within each error type of written work, small differences are observed. Students correctly identified the closest percentages of errors with a small difference of .6% in the evaluation of content and the furthest in the evaluation of vocabulary with a difference of .9%. A Mann-Whitney U test was used to determine whether any of these differences were significant.

4.5 RQ2: Statistical Analysis

A Mann-Whitney U test was conducted to determine whether the differences observed in the descriptive cross tables were significant. According to Nachar (2008), and McKnight and Najab (2010), to conduct a robust Mann-Whitney U test, the research data must meet four assumptions.

As has been discussed in the methodology chapter, two of the four assumptions have been met, which are a continuous dependent variable (from 0 to 100%) and two categorical independent variables (PA and SA). The third assumption is the independence of observations, which requires that there is no relationship between the observations in each group or between groups themselves. As argued in section 4.3, each student participant in the present study only participated in one and only one form of FA and had no chance to produce data under another form of FA; therefore, this assumption has been met.

The fourth assumption is that the two variables are not normally distributed. The quantitative data collected in the present study are transformed into split histograms to see if it satisfies this assumption.



THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

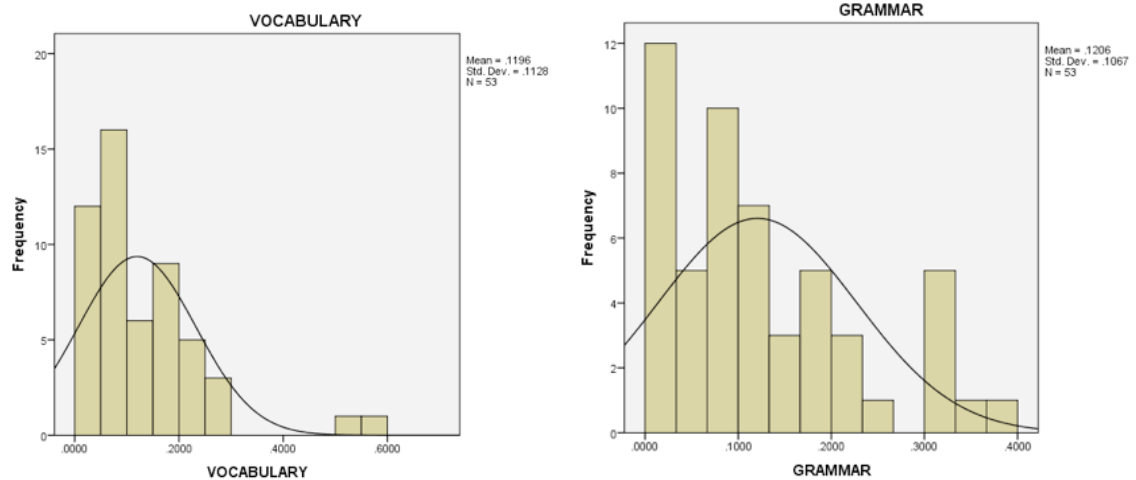


Figure 8. The split histograms of of SCIE/ OE total ratio

As shown in Figure 8, the four data sets used for the statistical analysis are not normally distributed; they have similar slightly skewed right shapes, which satisfies the last assumption for a robust Mann-Whitney U test.

Table 12

Mean ranks from Mann-Whitney U test

Error	Assessment	N	Mean Rank	Sum of Ranks
Types	Types			
Content	PA	26	26.83	697.50
	SA	27	27.17	733.50
Organization	PA	26	26.83	697.50
	SA	27	27.17	733.50
Vocabulary	PA	26	26.58	691.00
	SA	27	27.41	740.00
Grammar	PA	26	26.37	685.50

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

SA	27	27.61	745.50
----	----	-------	--------

Table 12 summarizes the ranks of means on four error types of written work. As shown by the sum of ranks, SA appears to be the type of assessment with the highest percentages overall. However, comparison within each assessment type indicates that students in the PA group corrected more errors related to content and organization while students in the SA group corrected more vocabulary and grammar errors.

Table 13

Mann-Whitney U test results

	Content	Organization	Vocabulary	Grammar
Mann-Whitney U	346.500	346.500	340.000	334.500
Wilcoxon W	697.500	697.500	691.000	685.500
Z	-.081	-.082	-.197	-.295
Asymp. Sig. (2-tailed)	.935	.935	.844	.768

Table 13 presents the results from the Mann-Whitney U test which examines whether there is a significant difference in the focuses of feedback produced by either SA or PA on four error types of written work. The results show that there is no statistically significant difference on any of the four error types, which means the focuses of feedback by students in the SA and PA groups were highly similar. Specifically, for content, there was no significant difference in the feedback produced through SA (Mdn =.091) and PA (Mdn =.083) conditions; $U=345.5$, $p=.935$. For organization, there was no significant difference in the feedback produced through SA (Mdn =.071) and PA (Mdn =.069) conditions; $U=345.5$, $p=.935$. For vocabulary, there was

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

no significant difference in the feedback produced through SA (Mdn =.100) and PA (Mdn =.089) conditions; $U=340$, $p=.844$. Finally, for grammar, there was no significant difference in the feedback produced through SA (Mdn =.105) and PA (Mdn =.083) conditions; $U=334.5$, $p=.768$.

4.6 Summary

This chapter presents the quantitative data organized by the relevance to the two research questions proposed in the methodology section. Both descriptive data and statistical analyses have been made to answer the research questions. The results will be further explained in the following chapter where connections to previous literature will be built.

Chapter 5: Discussion

5.1 Introduction

In this chapter, the two research questions proposed in the literature review will be answered with descriptive data and statistical analysis results from the previous chapter. Following that, the theoretical, methodological and empirical implications of the present study will be discussed. By establishing connections to previous literature, this chapter furthers the understanding of the nature of FA in the specific research context and situates the present study within the grand research agenda of L2 assessment proposed by Turner and Purpura (2016). Further on, the limitations of the current research will be outlined in chronological order, including those that occurred during the planning, implementation, and data collection stages of the study. Next, directions for future research will be presented. Finally, a summary of the chapter will conclude this section.

5.2 Findings

5.2.1 Findings of RQ1

RQ1: How much feedback do students give on written assignments during SA, PA and CA? Specifically, do students produce more feedback through CA than through SA or PA alone?

The descriptive data from Table 4 to 6 as well as the ANOVA test results in Table 8 provide answers to this question.

The overall discrepancy between SCIE and OE indicates that students in PA performed better at error identifying compared to students in SA and CA. When broken down by error types, it seems that PA functions better in the evaluation of content, organization, and vocabulary than the other two assessment approaches, while CA is more effective in the evaluation of

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

grammar compared to PA and SA. However, the ANOVA test results reveal that these differences are not statistically significant. Therefore, in this study of two Chinese EFL writing classrooms at the tertiary level, students did not produce more feedback through CA than through SA or PA alone.

Theoretically, with access to both self-reflection and peer interaction, students are expected to produce more feedback through CA; however, consistent with Chen et al. (2016), Lee and Coniam (2013), Vasu et al. (2016), and Walker (2015), feedback from peers is sometimes ignored in a teacher-centered learning environment since the teacher has been set as the authoritative figure. This might be the explanation for the lack of difference in students' feedback through different assessment approaches for formative purposes. In this study, students may not have felt that they were enough of an authority figure to provide as much feedback on their peers' written work as they were giving to themselves. Consequently, they gave predominantly positive comments on peers' writing, and the criticism was not adopted by their peers.

There are two other findings related to the first research question. First, it is clear that students were not able to correctly identify as many errors as the reviewer did through any form of FA. This general under-performance on FA tasks is inconsistent with the results from many studies set in the tertiary-level classrooms, which found that students are able to identify a similar number of errors as their teachers (Anderson & Palm, 2017; Black & Wiliam, 1998b; Baleghizadeh & Hajizadeh, 2014; Kilic, 2016; Liu & Li, 2014; Tsai & Chuang, 2013). However, according to Havnes et al. (2012), students feel less confident about the quality of formative feedback when FA is rarely used; FA feedback is considered more useful by students in science-oriented disciplines; female students tend to be more critical towards peer feedback compared to

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

the male counterpart. In the present study, FA was practiced by students for the first time; the discipline setting was tertiary-level EFL writing classroom; female students constitute more than 90% of the participants. All three conditions are contrary to what Havnes et al. (2012) identified as factors contributing to more positive adoption of feedback from FA.

The findings of this study do not mean that FA is not useful for this context. The formative function of PA and SA still can be realized through careful interpretation of the assessment feedback during follow-up instruction by the teachers. As noted by Black and Wiliam (1998a), Lam (2013), and Sambell et al. (2012), the formative function of assessment is not an inherent attribute of any particular assessing assessment practice, but needs to be realized through implementation, execution, and interpretation.

It is particularly interesting to note that at least one student successfully implemented SA as he/she correctly identified as many errors as the reviewer did. Although the reasons behind this are not revealed by the available data, it implies that even with limited L2 proficiency, students still have the potential to engage in effective SA. However, the collected data do not reveal what attributes the student who effectively performed SA has. Future studies could collect qualitative data of students' perception of FA practices to identify these attributes.

Second, the percentages of SCIE out of OE in each error type reveal students could in general notice a decent proportion ($M=56.34\%$) of errors in the evaluation of content, organization and vocabulary through FA but not in the evaluation of grammar. A substantial proportion ($M= 68.13\%$) of students failed to correctly identify a decent amount of grammatical errors in all three forms of FA. The differences, although not statistically significant, indicate that students' metalinguistic knowledge, even in a collective way, is not proficient enough for effective FA for grammar errors. In other words, grammar might be considered as a challenging

error type of written work for L2 students to evaluate, and according to Suzuki (2014), students' perception of the difficulty of assessment tasks is negatively correlated to their FA performance.

On the other hand, this contradicts the finding in Yoon and Lee (2013) that low to intermediate level students benefited more from FA in their study. In the present study, what is considered potentially beneficial for students' academic progress is the competence to correctly identify various types of errors in written work. This finding indicates that students' competence in identifying grammatical errors is not as strong as in identifying other types of errors.

The speculated reason for the difference between findings in this study and in Yoon and Lee's study might be the variance of participants' L2 proficiency. In Yoon and Lee's (2013) study, participants had low to high L2 proficiency; therefore, with random grouping, participants with high proficiency might have been placed in the same group with participants with low proficiency and taken the leading role in providing peer feedback. However, the same situation did not exist in the present study because all participants shared limited English proficiency.

5.2.2 Findings of RQ2

RQ2: What aspects of written assignments do students give feedback on within these different types of assessment? Specifically, which aspects of written assignments do students give more feedback on within SA? Which aspects of written assignments do students give more feedback on within PA?

The descriptive data from Table 9 to 11 as well as the Mann-Whitney U test results in Table 12 and 13 provide answers to this question.

The frequency counts of SCIE in PA and SA demonstrate that similar quantities of feedback have been produced in the evaluation of content, organization, vocabulary, and grammar. This finding is also confirmed by the percentages of SCE out of the total number of

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

SCE in the evaluation of four error types of students' written work. By taking the differences observed in the descriptive data into Mann-Whitney U test, it shows that PA generates overall more feedback on the evaluation of content and organization while SA generates overall more feedback on the evaluation of vocabulary and grammar. Further analysis indicates no statistically significant difference regarding the types of feedback produced through PA and SA.

The results indicate that in the present research context, students did not favor PA or SA in the evaluation of content, organization, vocabulary, and grammar. In other words, feedback generated through PA and SA had limited differences by the categorization of error types. This finding is consistent with Lau (2016) and Carnell (2016) who claimed that PA and SA are conceptually connected. The lack of a difference between the two FA forms might be because students' metacognitive processes of PA and SA are quite similar. They primarily relied on self-reflection and remained skeptical towards peer interaction and peer feedback.

However, the small differences within PA and SA indicate that different forms of FA might benefit students' written competence in different ways, which is consistent with Boud et al. (2015) as they argued that to boost the effectiveness of FA, it is important to adopt various assessment forms in classroom teaching.

As a finding that is related to the second research question, the percentages of SCIE out of OE were quite low (smaller than 13%) across all three groups in the evaluation of content, organization, vocabulary, and grammar, which demonstrate a rather ineffective implementation of PA and SA. This finding is an expansion on the findings in relation to research question one. It contradicts the claim made by Yu and Lee (2016) as the authors argue that students with low L2 proficiency are competent in giving constructive peer feedback. In the present study, students with intermediate English proficiency only spotted a small proportion of errors. The overall lack

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

of advanced proficiency might have hindered students' attempt to identify more errors as groups. However, it should be noted that giving constructive peer feedback is not the same as identifying errors (Brown and Harris, 2014; Brown et al., 2015). As a matter of fact, it is plausible for students to give highly constructive feedback without being able to describe exactly what the error is.

Taking the research context into consideration, it is perhaps unsurprising that students provided their peers with only limited feedback. According to Tan (2013), the effectiveness of FA approaches depends on the exposure time to assessment practices and "the realistic gap in students' learning that may be bridged within that time" (p. 3). In this study, the effectiveness of assessment approaches for formative purposes is reflected by students' competence of correctly identifying errors in written work. Student participants in the present study had only one 45-minute training session before engaging in the FA activity, and their English proficiency was only intermediate. Therefore, the training they had during the interactive workshop did not grant them sufficient assessment literacy to carry out effective FA, and their limited English proficiency was only strong enough for them to identify a small amount of errors in written work. It seems unreasonable to expect them to perform a most effective FA session.

In addition, the present study is not the only one that reveals that PA and SA might not be as effective as advocated. Suñol et al. (2016) noted that there is a high deviation between SA and TA in the context of a European university. Similarly, Jung (2016) also found that PA is not as reliable as TA in the context of an Asian university. Regardless of the difference in contexts, these studies, along with the present one, stress that the correct way to view this ineffectiveness is to consider using PA and SA as teaching pedagogies to cultivate students' awareness of cognitive gaps and their ability to self-regulate the learning process. This view is consistent with

Brown and Harris (2014) and Brown et al. (2015) as they pointed out that the formative use of FA has long-term benefits for students other than providing immediate revisions on their written work. With proper training of FA, students will be able to cultivate a strong sense of critical thinking.

5.3 Implications

The findings offer a fresh view on the nature of FA as a learning tool and a teaching pedagogy at the tertiary level. By exploring the combined SA and PA model, the present study examines the relationship between self-reflection and peer interaction in EFL writing classrooms in the context of a Chinese university. The findings also attempt to complete the understandings of PA and SA about the feedback produced by students. Also, the research study investigated the implementation of FA in a highly teacher-centered, form-focused and test-oriented learning environment, which reveals the practical issues for future adaption in similar contexts. The theoretical, methodological and empirical implications will be discussed in the following section.

5.3.1 Theoretical implications. At the theoretical level, by situating itself within the research agenda of second language assessment proposed by Turner and Purpura (2016), in which the scholars pointed out emerging areas in L2 assessment that require further exploration, this study furthers the understanding of CBA and consolidates the theoretical foundations of FA. As argued by Ubaque Casallas and Pinilla Castellanos (2016), FA should primarily serve as a tool for learning and not as a method of evaluation. Findings from the present study confirmed this statement as student participants were not able to identify as many errors as the reviewer in their written work. This indicates that in the present research context, it is not appropriate to exploit the summative value of FA; in other words, FA cannot serve as a replacement of TA.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

Therefore, the purpose of conducting FA should be raising students' awareness of the gap between their current level of learning and the expected learning objectives.

This corresponds to the claims of previous studies that the formative value of PA and SA surpasses the summative value (Carnell, 2016; Fazel, 2015; Dlosic et al. 2016; Lappin-Fortin & Rye, 2014; Mahlberg, 2015). Clearly, in the present research context, using the FA practices for evaluative purposes is not sensible as the reliability are not guaranteed; however, this does not mean students cannot learn from the evaluation itself. To be specific, the opportunities for students to learn are scattered all over the assessment process: by comparing the analytic rubrics against their written work, students can learn how to review assignments critically; by exchanging opinions on written work with peers, students can learn how to articulate their judgements with proper linguistic signals; by prompting questions to teachers during the assessment process, students can clarify any unclear interpretations they hold about the rating criteria. With facilitating learning as the core purpose of implementing PA and SA, incorporating FA practices as part of the curriculum in the present research context would be beneficial for students in the long term. With the focus being placed on the formative role played by PA and SA, it goes along with self-reflection and peer interaction in promoting learners' autonomy.

On the other hand, the formative use of PA and SA benefits both learners and instructors. This is particularly obvious when it comes to the acquisition of grammar. The grammatical errors correctly identified by students serve as a reminder of their insufficient syntax and morphology knowledge, while the grammatical errors incorrectly identified by students inform instructors of learners' misunderstanding of certain grammatical rules.

5.3.2 Methodological implications. At the methodological level, unlike most previous studies which have referred to the discrepancy between student ratings and instructor ratings as

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

the indicator for the effectiveness of FA, the present study adopted the discrepancy in frequency counts of feedback as the parameter to determine the quality of FA conducted by students.

Although both numerical ratings and detailed narratives are considered beneficial for students' learning in assessment (Cho & Park, 2014), as argued in the literature review chapter, there is a call for the research focus to shift from grading to feedback. The advantages of adopting this shift are three-fold.

First, compared to numerical ratings, which are assigned by students without a second confirmation, the descriptive feedback produced through PA, SA, and CA underwent a quality check conducted by teachers to ensure that only correctly identified errors were included, which increased the reliability of the evaluation results. Also, using these student-correctly-identified errors as the parameter exempted the influence of random guessing during the assessment process and therefore fundamentally increased the robustness of the data analyses.

Second, convergence between instructor ratings and student ratings only represents one side of the effectiveness of FA. However, students' description of identified errors through FA offers a more multifaceted representation of the effectiveness of FA. The discrepancy of identified errors from students and instructors not only reveals how effective formative PA and SA are in a certain research context but also generates valuable data on the proportions of errors correctly and incorrectly identified by students. Numerical ratings can hardly represent the cognitive process of evaluation; however, with descriptive feedback, the present study shows how students execute their linguistic repertoire to assess written work. Moreover, students' description of identified errors holds the potential for a qualitative analysis of discourse pattern in further research which will reveal the metacognitive process of conducting effective FA on a more in-depth level

Third, as has been discussed by Rea-Dickins (2013), the uptake of feedback is what makes assessment practices formative. The present study encourages students to organize their feedback in a descriptive manner which is directly linked to the revision of drafts. The feedback is a more directly linked to the formative nature of PA and SA than pure numerical scorings. To a large extent, the process of reflecting on the errors identified through PA and SA and articulating the feedback in a descriptive manner is also a manifestation of learning.

5.3.2 Empirical implications. At the empirical level, the exploration of FA in the context of a Chinese university reveals the following issues for the effective implementation of FA in similar contexts. The combination of PA and SA does not necessarily lead to more feedback in the present research context. As stated before, students overall intermediate L2 proficiency entails that the linguistic knowledge they have, even in a collective situation, may not be functional for correctly identifying a large number of errors in written work. For example, based on the researcher's observation of the FA process, when student participants engaged in the peer discussion about suspected errors, they experienced a difficult time convincing each other to reach unanimity. Since no one in a group had an advanced level of L2, the peer interaction only reinforced the partial understanding each member held. In practice, according to Zhao (2014), one possible solution is to increase the involvement of the teacher in FA. When mediation of conflicts is given upon request during PA and SA, students are more inclined to draw conclusions on arguments over identified errors.

Second, as indicated by the results, feedback from PA focused more on content and organization while feedback from SA focuses more on vocabulary and grammar. This difference implied that in the present research context, each FA practice functioned differently in the evaluation of students' written work. Therefore, teachers should consider using mixed

assessment practices to boost students' learning efficiency. Specifically, when the teaching objective is to raise students' awareness of the organization and content, it may be more useful to use PA; when the teaching objective is to enrich and enhance students' grammar and vocabulary, SA may be more appropriate.

5.4 Limitations

Restricted to limited time and resources, the present study inevitably suffers from methodological and empirical limitations. The curriculum in the chosen research venue, Jining Medical University, required the experimental intervention to take place over a short period so that the regular teaching objectives would not be disturbed. Therefore, the time arrangement was under strict scrutiny from the university administration. The FA training workshop for students and the FA session took place within three days. The repercussions due to this short research window are two-fold.

First, training is a crucial part of sound FA practices since experiential factors are related to evaluation performance. Empirical evidence has shown that sufficient and continuous training will significantly improve students' assessment performance (Boud, 2015; Liu & Li, 2014; Tsai & Chuang, 2013). However, with only one workshop session and one practicing model provided to participants, students' assessment literacy is not guaranteed. In other words, their competence to carry out, understand, and execute FA tasks on written work may not be at the peak. Second, the short research period also poses limitations on fully developing the FA process. Previous literature shows that involving students in the development of rating rubrics has benefits for both assessors and assessees in FA (Panadero, 2016). However, in the present study, the analytic rubrics were generated by the researcher and given to participants in the workshops. There are

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

several measures taken in the present study to minimize the influence of the absence of students' involvement in the development of analytics rubrics.

As discussed in the literature review chapter, effective FA has the potential to boost students' self-efficacy and learning motivation (Heritage, 2007). While most studies have reported positive attitude towards FA from students, in the context of Asia, several studies have reported mixed feelings about peer feedback from students (Vasu et al., 2016). In the present study, students were observed to be highly motivated and engaged with the FA process, as they were constantly asking probing questions during the interactive workshop and the FA process. However, their attitudes were neither recorded through quantitative data such as questionnaires nor qualitative data such as post-study interviews. The lack of data on students' attitudes leads to an incomplete picture of the effectiveness of FA.

Also, the collected data, descriptive in nature, were treated equally in data coding regardless of the wording. In other words, the value of feedback is only represented by the numbers of errors identified. It is possible that some students tend to be more ambiguous in describing identified errors while others are more used to give detailed description. In addition, the conceptual overlap between SA and PA (Carnell, 2016) might have blurred the distinction between these two assessment approaches. In other words, maybe PA has a lot in common with CA, because SA is practically subsumed within PA. Therefore, the "independent observation" assumption made in the study might not be well justified. Consequently, it compromises the strength of the data analysis in the present study.

Finally, although there is a shift from focusing on ratings to focusing on the feedback in the present research study, the FA process is not complete. As pointed out by Liu and Carless (2006), the learning component of FA is the uptake of feedback in the revisions of the written

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

work. However, students in the present study did not have the opportunity to compose second drafts. Therefore, there is no data indicating whether feedback from students is adopted or abandoned. Also, whether students' adoption of the feedback be influenced by their limited English proficiency is not known. For example, when their peers pointed out that their paper was missing a conclusion section, whether the assessed be able to compose a sound conclusion paragraph is not certain. Moreover, there is no post-test on the retention of the linguistic knowledge they might have gained from the assessment process. Practically speaking, the research value of this study overweighs its teaching value to the participants. The core notion of FA is that it fosters learning; the assessment process itself is a learning process, is a process of activating the linguistic knowledge student obtained from previous learning and try to execute them on a practical level.

There are several factors that are beyond the control of the researcher but might have influenced the research outcomes. For example, there is a significantly unbalanced gender ratio. As seen in the methodology section, the female participants in the study constitute for over 90% of the whole research population, which is predetermined by the student body admitted to the university. Whether it has influenced the peer interaction between the female majority student body and the male minority student body is not investigated. Although many researchers argued that students of different proficiency levels will benefit equally from the FA, there are also studies indicating that students of higher academic levels perform better at evaluating themselves and their peers.

Overall, these limitations have certain negative effects on the comprehensiveness of the present study, but the main research questions have been answered with sufficient data analyses.

The future research directions inspired by these limitations will be discussed in the following section.

5.5 Directions for Future Research

By reviewing the limitations in the present study, the previous section highlights several directions for future research. First, the expansion of examination of FA in different contexts is needed. Context plays a significant role in interpreting this study. Participants from the Jining Medical University have very limited exposure to FA in past EFL writing, and they have been studying in a teacher-centered, form-focused, and test-oriented environment. These conditions together constitute a very specific yet common educational environment for EFL classes at the tertiary level in China. In this context, the three FA forms, by comparing the discrepancy between feedback from student assessors and teacher assessors, is not particularly effective. This is consistent with some research studies at the tertiary level but contradictory to most research results from western universities. It sheds light on the necessity of exploring the localization of FA. To be specific, how should teachers adapt to different contexts when implementing FA as a teaching practice? In other words, what are the influential factors with regard to the effective implementation of FA in EFL writing classes at the tertiary level?

Second, since the importance of extended training sessions has been widely recognized, it would be worthwhile to examine the effects of continuous training sessions on students' assessment performance, especially for students with limited L2 proficiency. The present study has a one-off interactive workshop with a practicing model, which may not offer students enough time to acquire necessary assessment literacy to conduct effective FA. There should be follow-up examinations of students' assessment literacy development after each training sessions as a way to ensure that students are equipped with sufficient assessment skills before conducting FA as a

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

part of research studies. If we treat the FA process as a learning process, students should theoretically be able to learn linguistic knowledge from active involvement in evaluation. However, whether it is a practical method or not needs further investigation. In other words, to what extent can students learn from the constructive feedback given by their peers or generated by themselves? Moreover, during the extended training session, there should be enough time to engage students in the development of analytic rating rubrics. The extent to which students' involvement will influence their assessment performance is crucial for the effective implementation of FA.

Third, with the measurement of effectiveness shifted from rating scores to feedback, this study experimentally expanded the measurement of the effectiveness of FA. However, there are other factors that are not measured in the present study. For example, students' attitudes towards FA, their preferences between PA and SA, teachers' involvement in FA process. These factors require more diversified research methods. The present study uses only quantitative data from the reflective sheets as a way to represent the effectiveness of FA, while in fact, there could be discourse analysis on the descriptive feedback given by students which will reveal more in-depth features of feedback, such as the accuracy of description of errors and its influence on the adoption of feedback on future revisions. Besides, qualitative data from group interviews with students after FA sessions will certainly be a valuable source of information regarding students' attitudes towards FA; focus group discussions with participating teachers will possibly reveal more issues related to the practicality of implementing FA in EFL writing classrooms.

Last, as has been mentioned in the limitation section, the FA process implemented in the present study is not complete. In other words, the formative nature of the assessment is not fully realized. This is because that the feedback has not been put into future revisions; therefore, the

extent to which students directly benefited from participating in the research study is not shown. In future research, there should be a post-test on knowledge retention, which investigates students' uptake of the feedback from FA and how students include the explicit feedback in the revision process. Although they have been exposed to feedback from either their own reflection or through peer discussion, the present study does not reveal whether they were capable of processing the feedback into sound revisions on their written work. What's more, since there are falsely identified errors, how teachers could make use of these errors to reinforce metalinguistic knowledge and stylistic conventions will be an intriguing topic.

5.6 Summary

This chapter discusses the data analysis from the results section in response to the two research questions proposed in the literature review chapter. Although statistical analyses revealed no significant differences across different groups, the present study still has significant implications as a research project conducted in a special educational context. The theoretical, methodological and empirical implications are discussed extensively, followed by a discussion of the limitations perceived during the research process. Finally, directions for future research are presented in response to the limitations. The next chapter summarizes the highlight points of the present study and recaps the significance for second language assessment.

Chapter 6. Conclusion

The present study corresponds to the call for enabling excellence in a changing research landscape proposed by Social Sciences and Humanities Research Council of Canada (SSHRC) through the SSHRC Strategic Plan to 2020: Advancing Knowledge for Canada's Future (2016). Increased international collaboration brings a growing body of emerging researchers with international backgrounds to Canada (Canadian Bureau for International Education, 2016). Undoubtedly, their innovation and creativity contribute to the advancement of the Canadian society. However, their diversified learning experiences and expectations pose new challenges to Canadian higher education. This is especially true in the context of higher education. As the training provided to students mainly focus on enhancing research skills needed for future academic progress.

By exploring FA practices in other contexts, the present study will facilitate Canadian higher educational institutions to understand better the language challenges and barriers facing these emerging researchers and configure a set of effective measures to improve their academic written competence to make their achievements more comprehensible and accessible to Canadians in an evolving society. The present research contributes to promoting and supporting the knowledge in the aspects of creation, accessibility, and mobilization.

As one of the pioneering studies exploring the effectiveness of PA and SA in tertiary level EFL writing classes in China, the present study sheds light on the merits and challenges facing the implementation of FA in a teacher-centered, test-oriented, text-book-analysis based learning environment. Building on previous literature, the findings of this study further prove that FA should not only be viewed as an evaluation tool but also a powerful engine promoting learning as well as an informative reference facilitating instruction.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

By attempting to answer the proposed research questions with analyses of quantified feedback from PA and SA, the present study reveals that integrating PA and SA do not necessarily lead to more feedback being provided by students; also, the feedback students gave through PA and SA do not differ significantly in the sense of quantity. In other words, the combination of PA and SA, in the present research context, has no obvious advantages over PA or SA conducted separately; students performed quite equally on PA and SA as reflected by the feedback they gave. It is likely that students were not able to perform effective FA on written work due to limited L2 proficiency and insufficient prior exposure to FA practices; however, the feedback they gave by conducting PA and SA has slightly different focuses in the evaluation of content, organization, vocabulary, and grammar. Specifically, students performed better at identifying contextual and organizational errors with PA, and lexical and syntactic errors with SA.

With these findings, the study provides implications for research and practice in similar contexts. First and foremost, the shift from using numerical ratings to using the quantity of feedback as the primary parameter to determine the quality of assessment practices explores a different way of conducting quantitative studies in second language assessment; secondly, possible ways of improving students' performance in FA have been proposed, including extending the length of training sessions, increasing instructors' involvement during FA, and pairing students with different L2 proficiencies.

It is crucial to point out that context is a major component in the present study. It has significant impacts on the research process as well as the outcomes. Although not all influential factors are identified through the collected data, by comparing the findings in the current study with research studies conducted in similar contexts, it is obvious that participants' demographic

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

backgrounds and the overall educational setting should be considered in the interpretation of the research findings. In the same sense, the generalization of research findings from the present study should always be carried out with prudence.

Reference

- Afitska, O. (2014). Use Of Formative Assessment, Self-And Peer-Assessment In The Classrooms: Some Insights From Recent Language Testing And Assessment (LTA) Research. *i-Manager's Journal on English Language Teaching*, 4(1), 29.
- Allen, D., & Mills, A. (2016). The impact of second language proficiency in dyadic peer feedback. *Language Teaching Research*, 20(4), 498-513.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

- Ashenafi, M. M. (2017). Peer-assessment in higher education—twenty-first century practices, challenges and the way forward. *Assessment & Evaluation in Higher Education*, 42(2), 226-251.
- Andersson, C., & Palm, T. (2017). The impact of formative assessment on student achievement: a study of the effects of changes to classroom practice after a comprehensive professional development programme. *Learning and Instruction*, 49, 92-102.
- Anderson, J. N. (2012). Student involvement in assessment: Healthy self-assessment and effective peer assessment. *The Cambridge guide to assessment*, 187-197.
- Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self assessment. *Theory into practice*, 48(1), 12-19.
- Andrade, H. L. (2010). Students as the definitive source of formative assessment. *Handbook of formative assessment*, 90-105.
- Assessment Reform Group. (1999). Assessment for learning: Beyond the black box. *Université de Cambridge, School of Education*.
- Baleghizadeh, S., & Hajizadeh, T. (2014). Self-and Teacher-Assessment in an EFL Writing Class. *Gist: Education and Learning Research Journal*, (8), 99-117.
- Bachman, L. F., & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.
- Baleni, Z. G. (2015). Online Formative Assessment in Higher Education: Its Pros and Cons. *Electronic Journal of e-Learning*, 13(4), 228-236.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

- Bennett, R. E., & Gitomer, D. H. (2009). Transforming K–12 assessment: Integrating accountability testing, formative assessment and professional support. In *Educational assessment in the 21st century* (pp. 43-61). Springer Netherlands.
- Black, H. (1986). Assessment for learning. In D. L. Nuttall (Ed.), *Assessing Educational Achievement*. (pp. 7-18). London: Falmer Press.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5(1), 7-74.
- Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2).
- Black, P., & Wiliam, D. (2003). In Praise of Educational Research: Formative Assessment. *British Educational Research Journal* 29 (5): 623–637.
- Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. *Educational evaluation: new roles, new means: the 63rd yearbook of the National Society for the Study of Education*, (part II), 26-50.
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4), 399-413.
- Boud, D., Lawson, R., & Thompson, D. G. (2015). The calibration of student judgement through self-assessment: disruptive effects of assessment patterns. *Higher Education Research & Development*, 34(1), 45-59.
- Broadfoot, P. M., Daugherty, R., Gardner, J., Harlen, W., James, M., & Stobart, G. (2002). *Assessment for learning: 10 principles*. Cambridge, UK: University of Cambridge School of Education.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

- Brown, G. T., Andrade, H. L., & Chen, F. (2015). Accuracy in student self-assessment: directions and cautions for research. *Assessment in Education: Principles, Policy & Practice*, 22(4), 444-457.
- Brown, G. T. L., & Harris, L. R. (2013a). Student self-assessment. In J. H. McMillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 367–393). Thousand Oaks: Sage.
- Brown, G. T. & Harris, L. R. (2013b). Opportunities and obstacles to consider when using peer and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education*, 36, 101-111.
- Brown, G. T. L., & Harris, L. R. (2014). The future of self-assessment in classroom practice: Reframing self-assessment as a core competency. *Frontline Learning Research*, 2(1), 22-30.
- Canadian Bureau for International Education (2016). Canada's performance and potential in international education. Retrieved from: <http://cbie.ca/media/facts-and-figures/>
- Carless, D. (2007). Learning-oriented assessment: conceptual bases and practical implications. *Innovations in Education and Teaching International*, 44(1), 57-66.
- Carless, D. (2012). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings*. Routledge.
- Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education*, 69(6), 963.
- Carnell, B. (2016). Aiming for autonomy: formative peer assessment in a final-year undergraduate course. *Assessment & Evaluation in Higher Education*, 41(8), 1269-1283.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

- Cheng, L. (2008). The key to success: English language testing in China. *Language Testing*, 25(1), 15-37.
- Chen, Q., May, L., Klenowski, V., & Kettle, M. (2014). The enactment of formative assessment in English language classrooms in two Chinese universities: Teacher and student responses. *Assessment in Education: Principles, Policy & Practice*, 21(3), 271-285.
- Cho, Y. H., & Cho, K. (2011). Peer reviewers learn from giving comments. *Instructional Science*, 39(5), 629-643.
- Cho, S., & Park, C. (2014). The Role of Scoring In Formative Assessment of Second Language Writing. *GEMA Online® Journal of Language Studies*, 14(3).
- Cizek, G. J. (2010). An introduction to formative assessment. In Andrade, H., & Cizek, G. J. (Eds.), *Handbook of formative assessment* (pp. 3-17). New York: Routledge.
- Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, 24(2), 205-249.
- Cohen, L., Manion, L., & Morrison, K. (2013). *Research methods in education*. Routledge.
- Crusan, D. (2010). *Assessment in the second language writing classroom*. University of Michigan Press.
- Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *Tesol Quarterly*, 393-415.
- Dolotic, H. N., Brantmeier, C., Strube, M., & Hoglebe, M. C. (2016). Living Language: Self Assessment, Oral Production, and Domestic Immersion. *Foreign Language Annals*, 49(2), 302-316.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

- Ertle, B., Rosenfeld, D., Presser, A. L., & Goldstein, M. (2016). Preparing preschool teachers to use and benefit from formative assessment: the Birthday Party Assessment professional development system. *Zentralblatt für Didaktik der Mathematik*, 7(48), 977-989.
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta analysis. *Review of Educational Research*, 59(4), 395-430.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta analysis comparing peer and teacher marks. *Review of educational research*, 70(3), 287-322.
- Fazel, I. (2015). A Step in the Right Direction: Peer-assessment of Oral Presentations in an EFL Setting. *Reading Matrix: An International Online Journal*, 15(1), 78-90.
- Fishbach, A. & Eyal, T., & Finkelstein, S. R. (2010). How positive and negative feedback motivate goal pursuit. *Social and Personality Psychology Compass*, 4, 517-530.
- Gardner, J. N., & Gardner, J. (Eds.). (2012). *Assessment and learning*. Sage.
- Germaine, R., Richards, J., Koeller, M., & Schubert-Irastorza, C. (2016). Purposeful Use of 21st Century Skills in Higher Education. *Journal of Research in Innovative Teaching*, 9(1).
- Goodrich, H. (1996). *Student self-assessment: At the intersection of metacognition and authentic assessment* (Doctoral dissertation, Harvard Graduate School of Education).
- Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta analysis. *The Elementary School Journal*, 115(4), 523-547.
- Havnes, A., Smith, K., Dyshe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation*, 38(1), 21-27.
- Heritage, M. (2007). Formative assessment: What do teachers need to know and do?. *Phi Delta Kappan*, 89(2), 140-145.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

- Hill, K., & McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3), 395-420.
- Huang, S. C. (2016). Understanding learners' self-assessment and self-feedback on their foreign language speaking performance. *Assessment & Evaluation in Higher Education*, 41(6), 803-820.
- Hunter, D., Mayenga, C., & Gambell, T. (2006). Classroom assessment tools and uses: Canadian English teachers' practices for writing. *Assessing Writing*, 11(1), 42-65.
- James, M. (2006). Assessment, teaching and theories of learning. *Assessment and learning*, 47-60.
- Jiang, X. (2013). Challenges for college-level learners of academic English writing in China.
- Jin, Y., & Fan, J. (2011). Test for English majors (TEM) in China. *Language Testing*, 28(4), 589-596.
- Jung, M. Y. (2016). Peer/Teacher-Assessment Using Criteria in the EFL Classroom for Developing Students L2 Writing. *Pan-Pacific Association of Applied Linguistics (PAAL)*, 20(1), 1-20.
- Ketabi, S., & Ketabi, S. (2014). Classroom and formative assessment in second/foreign language teaching and learning. *Theory and Practice in Language Studies*, 4(2), 435.
- Kilic, D. (2016). An Examination of Using Self-, Peer-, and Teacher-Assessment in Higher Education: A Case Study in Teacher Education. *Higher Education Studies*, 6(1), 136-144.
- Kulik, J. A., & Kulik, C. L. C. (1988). Timing of feedback and verbal learning. *Review of educational research*, 58(1), 79-97.
- Lam, A. S. (2005). *Language education in China: Policy and experience from 1949: Policy and experience from 1949* (Vol. 1). Hong Kong University Press.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

- Lam, R. 2013. "Formative Use of Summative Tests: Using Test Preparation to Promote Performance and Self-regulation." *Asia-Pacific Educational Researcher* 22 (1): 69–78.
- Lappin-Fortin, K., & Rye, B. J. (2014). The Use of Pre-/Posttest and Self-Assessment Tools in a French Pronunciation Course. *Foreign Language Annals*, 47(2), 300-320.
- Lau, A. M. S. (2016). 'Formative good, summative bad?'—A review of the dichotomy in assessment literature. *Journal of Further and Higher Education*, 40(4), 509-525.
- Lee, I., & Coniam, D. (2013). Introducing assessment for learning for EFL writing in an assessment of learning examination-driven system in Hong Kong. *Journal of Second Language Writing*, 22(1), 34-50.
- Leung, C. (2004). Developing formative teacher assessment: Knowledge, practice, and change. *Language Assessment Quarterly: An International Journal*, 1(1), 19-41.
- Leung, C., & Lewkowicz, J. (2006). Expanding horizons and unresolved conundrums: Language testing and assessment. *TESOL Quarterly*, 211-234.
- Li, H., Xiong, Y., Zang, X., L. Kornhaber, M., Lyu, Y., Chung, K. S., & K. Suen, H. (2016). Peer assessment in the digital age: a meta-analysis comparing peer and teacher ratings. *Assessment & Evaluation in Higher Education*, 41(2), 245-264.
- Liu, N. F., & Carless, D. (2006). Peer feedback: the learning element of peer assessment. *Teaching in Higher education*, 11(3), 279-290.
- Liu, X., & Li, L. (2014). Assessment training effects on student assessment skills and task performance in a technology-facilitated peer assessment. *Assessment & Evaluation in Higher Education*, 39(3), 275-292.
- Lombard, B. J. J. (2008). Modelling critical thinking through learning-oriented assessment. *South African Journal of higher education*, 22(1), 1029-1043.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

- McKnight, P. E., & Najab, J. (2010). Mann-Whitney U Test. *Corsini Encyclopedia of Psychology*.
- Mahlberg, J. (2015). Formative self-assessment college classes improves self-regulation and retention in first/second year community college students. *Community College Journal of Research and Practice*, 39(8), 772-783.
- Mattisson Ekstam, J. (2015). Cultural and academic meetings in the writing classroom: China and the West. *Sino-US English Teaching*, 12(4), 241-249.
- McMillan, J. H., Cohen, J., Abrams, L., Cauley, K., Pannozzo, G., & Hearn, J. (2010). Understanding Secondary Teachers' Formative Assessment Practices and Their Relationship to Student Motivation. *Online Submission*.
- NACFLT (2007). *Syllabus for University English Language Teaching*. Shanghai: Shanghai Foreign Language Education Press.
- NACFLT (2015). *Syllabus for TEM4*. Shanghai: Shanghai Foreign Language Education Press.
- Nachar, N. (2008). The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution. *Tutorials in Quantitative Methods for Psychology*, 4(1), 13-20.
- Netten, J. E., & Spain, W. H. (1989). Student-Teacher Interaction Patterns in the French Immersion Classroom: Implications for Levels of Achievement in French Language Proficiency. *Canadian Modern Language Review*, 45(3), 485-501.
- Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education*, 39(1), 102-122.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

- Nielsen, K. (2014). Self-assessment methods in writing instruction: a conceptual framework, successful practices and essential strategies. *Journal of Research in Reading*, 37(1), 1-16.
- O'Donnell, A. M., & Topping, K. (1998). Peers assessing peers: Possibilities and problems. *Peer-assisted learning*, 255-278.
- O'Sullivan, B. (2012). A brief history of language testing. *The Cambridge guide to second language assessment*, 9-19.
- Panadero, E., Brown, G. T., & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational psychology review*, 28(4), 803-830.
- Purpura, J. (2004). Learning-oriented assessment for grammatical ability. *Assessing Grammar*, 212-250.
- Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal*, 100(S1), 190-208.
- Rashid, R. A., & Jaidin, J. H. (2014). Exploring Primary School Teachers' Conceptions of "Assessment for Learning". *International Education Studies*, 7(9), 69.
- Rea-Dickins, P. (2002). Exploring the educational potential of assessment with reference to learners with English as an additional language. *Language and additional/second language issues for school education: A reader for teachers*, 81-93.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.
- Sambell, K., L. McDowell, and C. Montgomery. 2012. *Assessment for Learning in Higher Education*. Abingdon: Routledge.
- Sato, M., & Ballinger, S. (2016). Peer Interaction and Second Language Learning.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

- Scriven, M. S. (1967). The methodology of evaluation (Perspectives of Curriculum Evaluation, and AERA monograph Series on Curriculum Evaluation, No. 1). *Chicago: Rand McNally*.
- Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, 78(1), 153-189.
- Stoynoff, S. (2012). Looking backward and forward at classroom-based language assessment. *ELT journal*, 66(4), 523-532.
- Social Sciences and Humanities Research Council of Canada (2016). SSHRC strategic plan to 2020: Advancing knowledge for Canada's future. Retrieved from: http://www.sshrc.gc.ca/about-au_sujet/publications/strategic-plan-strategique-2016-eng.pdf
- Song, Y. (2011). *Assessment for Learning in a Chinese University Context: A Mixed Methods Case Study on English as a Foreign Language Speaking Ability* (Doctoral dissertation, McGill University).
- Strong-Krause, D. (2000). Exploring the effectiveness of self-assessment strategies in ESL placement. *Learner-directed assessment in ESL*, 49-73.
- Suen, H. K. (2014). Peer assessment for massive open online courses (MOOCs). *The International Review of Research in Open and Distributed Learning*, 15(3).
- Suñol, J. J., Arbat, G., Pujol, J., Feliu, L., Fraguell, R. M., & Planas-Lladó, A. (2016). Peer and self-assessment applied to oral presentations from a multidisciplinary perspective. *Assessment & Evaluation in Higher Education*, 41(4), 622-637.
- Suzuki, Y. (2015). Self-assessment of Japanese as a second language: The role of experiences in the naturalistic acquisition. *Language Testing*, 32(1), 63-81.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

- Tan, K. (2013). A Framework for Assessment for Learning: Implications for Feedback Practices within and beyond the Gap. *ISRN Education*, 2013.
- Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of educational Research*, 68(3), 249-276.
- Topping, K. J. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In *Optimising new modes of assessment: In search of qualities and standards* (pp. 55-87). Springer Netherlands.
- Topping, K. J. (2009). Peer assessment. *Theory into practice*, 48(1), 20-27.
- Topping, K. J. (2010a). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction*, 20(4), 339-343.
- Topping, K. J. (2010b). Peers as a source of formative assessment. *Handbook of formative assessment*, 61-74.
- Torrance, H. (1993). Formative assessment: Some theoretical problems and empirical questions. *Cambridge Journal of Education*, 23(3), 333-343.
- Tsai, Y. C., & Chuang, M. T. (2013). Fostering revision of argumentative writing through structured peer assessment. *Perceptual and motor skills*, 116(1), 210-221.
- Turner, E. C. (2013). Classroom Assessment. In Fulcher, G. & Davidson, F. (Eds.), *The Routledge Handbook of Language Testing*. London and New York: Routledge.
- Turner, J. R., & Thayer, J. (2001). *Introduction to Analysis of Variance: Design, Analysis & Interpretation: Design, Analysis & Interpretation*. Sage.
- Turner, E. C. & Purpura, J. E. (2016). Learning oriented assessment in second and foreign language classrooms. In Tsagari D. & Baneerjee J. (Eds.), *Handbook of Second Language Assessment* (pp. 255-272). DeGruyter, Inc.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

- Ubaque Casallas, D. F., & Pinilla Castellanos, F. S. (2016). Argumentation Skills: A Peer Assessment Approach to Discussions in the EFL Classroom. *Profile Issues in TeachersProfessional Development*, 18(2), 111-123.
- Ünaldı, İ. (2016). Self and teacher assessment as predictors of proficiency levels of Turkish EFL learners. *Assessment & Evaluation in Higher Education*, 41(1), 67-80.
- Vanderlelie, J. J., & Alexander, H. (2016). Learning-oriented assessment increases performance and written skills in a second year metabolic biochemistry course. *Biochemistry and Molecular Biology Education*, 44(4), 412-420.
- Van Zundert, M., Sluijsmans, D., & Van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, 20(4), 270-279.
- Vasu, K., Ling, C. H., & Nimehchisalem, V. (2016). Malaysian Tertiary Level ESL Students' Perceptions toward Teacher Feedback, Peer Feedback and Self-assessment in their Writing. *International Journal of Applied Linguistics and English Literature*, 5(5), 158-170.
- Walker, M. (2015). The quality of written peer feedback on undergraduates' draft answers to an assignment, and the use made of the feedback. *Assessment & Evaluation in Higher Education*, 40(2), 232-247.
- Way, W. D., Dolan, R. P., & Nichols, P. A. U. L. (2010). Psychometric challenges and opportunities in implementing formative assessment. *Handbook of formative assessment*, 297-315.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

- Weurlander, M., Söderberg, M., Scheja, M., Hult, H., & Wernerson, A. (2012). Exploring formative assessment as a tool for learning: students' experiences of different methods of formative assessment. *Assessment & Evaluation in Higher Education*, 37(6), 747-760.
- Wiliam, D. (2011). What is assessment for learning?. *Studies in Educational Evaluation*, 37(1), 3-14.
- Wiliam, D. (2013). Assessment: The bridge between teaching and learning. *Voices from the Middle*, 21(2), 15.
- Wiliam, D. (2014, April). Formative assessment and contingency in the regulation of learning processes. In *Annual Meeting of American Educational Research Association*, Philadelphia, PA.
- Wingate, U. (2010). The impact of formative feedback on the development of academic writing. *Assessment & Evaluation in Higher Education*, 35(5), 519-533.
- Winne, P. H., & Butler, D. L. (1994). Student cognition in learning from teaching. In T. Husen & T.
- Pedhazur, E. J., & Schmelkin, L. P. (2013). *Measurement, design, and analysis: An integrated approach*. Psychology Press.
- Postlewaite (Eds.), *International encyclopedia of education* (2nd ed., pp. 5738–5745). Oxford: Pergamon.
- Yang, Y. (2000). History of English Education in China (1919-1998).
- Yoon, E., & Lee, H. K. (2013). Do effects of self-assessment differ by L2 language level? A case of Korean learners of English. *The Asia-Pacific Education Researcher*, 22(4), 731-739.
- Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher education*, 45(4), 477-501.

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

- Yu, S., & Lee, I. (2016). Understanding the Role of Learners With Low English Language Proficiency in Peer Feedback of Second Language Writing. *TESOL Quarterly*, 50(2), 483-494.
- Zhao, H. (2014). Investigating teacher-supported peer assessment for EFL writing. *ELT journal*, 68(2), 155-168.

Appendix A: Initial Communication



McGill

Faculty of
Education

Department of
Integrated Studies in Education

Dear University Administrator,

I'm writing to obtain your approval of conducting my research in your university. My research is designed for a thesis that will be submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Arts in Second Language Education. The topic of my research is the effectiveness of formative assessment at the tertiary level. Your participation

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

will involve approximately 40 students who major in English attending a workshop introducing formative assessment and a regular academic writing class session. In the writing class, students will conduct formative assessment (self and/or peer assessment) on their own written work. In conjunction with students' voluntary participation in this study, three teachers are needed for reviewing students' written work and giving feedback.

The purpose of this study is to understand how self and peer assessment work in a separated and combined way, and to investigate the differential effect of feedback produced through self and peer assessment respectively. The benefits of this study extend from student participants to teacher participants. Students will learn the effective way to assess their own and their peers' work and to produce constructive feedback; while teachers will understand how self and peer assessment interact with each other in second language writing classroom and have a clearer direction in future class planning.

As in any kind of research of this kind, no one will be identified by name and so I can ensure confidentiality. The data collected from this study will be stored on my personal computer in password-protected files.

If you gave any questions about this research or would like to withdraw your consent at any time, please feel free to contact me at 438-926-4291 or by email at mo.guo@mail.mcgill.ca. If you have any ethical concerns or complaints about your participation in this study, and want to speak with someone not on the research team, please contact the McGill Ethics Manager at 514-398-6831 or lynda.mcneil@mcgill.ca.

Sincerely,

Mo Guo

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

MA candidate of Second Language Education

Department of Integrated Studies in Education

McGill University

Montreal, QC, Canada

Appendix B: Student Consent Form



Participant Consent Form

McGill University
Faculty of Education

Department of Integrated Studies of Education

Researcher: Mo Guo
MA candidate
McGill University
Department of Integrated Studies in Education
438-926-4291
mo.guo@mail.mcgill.ca

Supervisor: Dr. Susan Ballinger
Assistant Professor
McGill University
Department of Integrated Studies in Education
514-769-1906
susan.ballinger@mcgill.ca

Title of Project: The Effectiveness of Formative Assessment in the Context of a Chinese University

Dear Student:

You are invited to participate in a research study about the use of self and peer assessment at university level. Self and peer assessment have been used in second language learning classroom for a long time. Their usefulness in second language writing classroom has been

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

confirmed by many previous studies. However, these studies only looked into the use of either self or peer assessment at university level, even though both forms are useful. Therefore, the purpose of this study is to see how self and peer assessment work when used in a separate and a combined form in second language writing classrooms.

Your participation will involve attendance at your next academic writing class, during which you will be asked to write an essay on a given topic. Then you will be randomly assigned to one of several groups and you will conduct self and/or peer assessment on your and/or your peers' essays. Lastly, you will transcribe the feedback you produce or receive from the assessment process onto a reflective sheet distributed in class. The approximate time commitment is 90 minutes, the same length of one regular class session.

Your participation in this project is voluntary. You have the right to refuse to participate in parts of the study, to decline to answer any question, or to withdrawal from the study at any time, for any reason. Whether you choose to participate or not has no influence on your class standing or test grades. You will not be penalized for withdrawing from the study. If you decide to withdraw, your written assignments will be removed from the research files and returned to you. If you choose not to participate in the study, or withdraws from the study, this information will not be made available to the school administrators. There are no anticipated risks to you by participating in this research.

By participating in this research, you will be equipped with the knowledge of how to engage in self and peer assessment, and you will receive detailed feedback on your written assignments. Moreover, you will benefit directly from this study as it can help you in preparing for the Test for English Majors Level 4 (TEM4). During the research period, you will have the opportunity to learn effective peer communicative skills and practice conducting objective and constructive

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

formative assessment. As academic writing is a big portion of your compulsory English level tests, you will have an opportunity to practice and improve your written English.

No personal identifying information is to be shown in the final publication. The data to be collected in this study is the number of errors identified in your sample essay. Your name will be coded into a combination of numbers and capital English letters. All the research data will be stored on the researcher's personal computer in password-protected files. Only the research and the faculty supervisor will have access to identifiable data.

If you have any questions about this research or would like to withdraw your consent at any time, please feel free to contact me at 438-926-4291 or by email at mo.guo@mail.mcgill.ca. If you have any ethical concerns or complaints about your participation in this study, and want to speak with someone not on the research team, please contact the McGill Ethics Manager with REB file # 125-0816 at 514-398-6831 or lynda.mcneil@mcgill.ca.

Please sign below if you have read the above information and consent to participate in this study. Agreeing to participate in this study does not waive any of your rights or release the researchers from their responsibilities. A copy of this consent form will be given to you and the researcher will keep a copy.

Participant's Name (please print): _____

Participant's Signature: _____ Date: _____

Appendix C: Teacher Consent Form



Participant Consent Form

McGill University
Faculty of Education
Department of Integrated Studies of Education

Researcher: Mo Guo
MA candidate
McGill University
Department of Integrated Studies in Education
438-926-4291
mo.guo@mail.mcgill.ca

Supervisor: Dr. Susan Ballinger
Assistant Professor
McGill University
Department of Integrated Studies in Education
514-769-1906
susan.ballinger@mcgill.ca

Title of Project: The Effectiveness of Formative Assessment in the Context of a Chinese University

Dear Teacher:

You are invited to participate in a research study about the use of self- and peer assessment at the tertiary level. Self- and peer assessment have been used in second language learning classroom for a long time. Their usefulness in second language writing classrooms has been confirmed by many previous studies. However, these studies only looked into the use of either self or peer assessment at university level, even though both forms are useful. Therefore, the purpose of this study is to see how self and peer assessment work when used in a separate and a combined form in second language writing classroom.

Your participation of this research will involve attendance at the next academic writing class. During the academic writing class, student participants will be asked to write an essay on a given topic. Then they will be randomly assigned to one of several groups and conduct self and/or peer assessment on their and/or their peers' essays. Lastly, participants will transcribe the feedback they produce or receive from the

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

assessment process onto a reflective sheet distributed in class. After that, one teacher will evaluate students' feedback on their written work and two other teachers will evaluate students' essays based on an analytic rubrics and transcribe the feedback onto a reflective sheet. The total approximate time commitment is 180 minutes, about the same length of two regular class sessions.

Your participation in this project is voluntary. You have the right to refuse to participate in parts of the study, to decline to answer any question, or to withdraw from the study at any time, for any reason. Whether you choose to participate or not has no influence on your job performance or teacher evaluation. You will not be penalized for withdrawing from the study. There are no anticipated risks to you by participating in this research.

By participating in this research, you will be equipped with the knowledge of engaging students in formative assessment and how to provide detailed feedback on students' written assignments. During the research period, you will have the opportunity to learn effective classroom management skills and practice conducting objective and constructive formative assessment. As academic writing is a big portion of participants' compulsory English level tests, your students will have an opportunity to practice and improve their written English and thereof reducing your workload.

No personal identifying information is to be shown in the final publication. The data to be collected in this study is the number of mistakes identified in students' sample essays. Their name will be coded into a combination of numbers and capital English letters. All the research data will be stored on the researcher's personal computer in password-protected files. Only the research and the faculty supervisor will have access to identifiable data.

If you have any questions about this research or would like to withdraw your consent at any time, please feel free to contact me at 438-926-4291 or by email at mo.guo@mail.mcgill.ca. If you have any ethical concerns or complaints about your participation in this study, and want to speak with someone not

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

on the research team, please contact the McGill Ethics Manager with REB file #125-0816 at 514-398-6831 or lynda.mcneil@mcgill.ca.

Please sign below if you have read the above information and consent to participate in this study. Agreeing to participate in this study does not waive any of your rights or release the researchers from their responsibilities. A copy of this consent form will be given to you and the researcher will keep a copy.

Participant's Name: (please print) _____

Participant's Signature: _____ Date: _____

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

Appendix D: Analytic Rubrics

Writer:

Evaluator:

Content							Points
Sufficiently summarizes the main message of the excerpt	5	4	3	2	1	0	
Includes relevant comments on the main message	5	4	3	2	1	0	
Uses multiple paragraphs to successfully support an argument	5	4	3	2	1	0	
Uses specific details and examples from personal knowledge	5	4	3	2	1	0	
Uses sentences with clear and coherent meanings	5	4	3	2	1	0	
Organization							
Has a clear introduction, body and conclusion format	5	4	3	2	1	0	
Uses transitions and other logical connectors appropriately	5	4	3	2	1	0	
Uses appropriate general-specific organization	5	4	3	2	1	0	
Has a clear topic sentence that is easy to find (at end of intro)	5	4	3	2	1	0	
Includes concluding sentences that are effectively summative	5	4	3	2	1	0	
Vocabulary							
Uses a range of varied vocabulary	5	4	3	2	1	0	
Has appropriate word choice and form that does not affect meaning	5	4	3	2	1	0	
Shows knowledge of correct spelling	5	4	3	2	1	0	
Uses effective punctuation	5	4	3	2	1	0	
Grammar							
Verbs: correctly uses “be” verbs, tense choices, subjects and verbs agree	5	4	3	2	1	0	

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

Nouns, articles and quantifiers: correctly uses singular and plural forms, articles (a, an, the) and quantifiers (some, any, many, much, a few, a little and a lot of etc.)	5	4	3	2	1	0	
Connectors: correctly uses and, but, so, because etc.	5	4	3	2	1	0	
Uses a variety of sentence structures (complex, simple, compound)	5	4	3	2	1	0	
Avoids run-ons (fused sentence) and comma splices	5	4	3	2	1	0	
Avoids incomplete sentences (fragments)	5	4	3	2	1	0	

Appendix E: TEM4 Writing Prompt

Writing Prompt (45 MIN)

*Read carefully the following excerpt on holidays during school time discussion in the UK, and then write your response in **NO LESS THAN** 200 words, in which you should:*

- *Summarize the main message of the excerpt, and then*
- *Comment on whether parents should take children out of school for holiday during term time in order to save money.*

You should support yourself with information from the excerpt.

Marks will be awarded for content relevance, content sufficiency, organization and language quality. Failure to follow the above instructions may result in a loss of marks.

No More Holidays during School Time

Parents will not be allowed to take children out of school for family vacations any more, according to UK's Education Secretary, Michael Gove.

He will take away the right of head teachers to "permit absence" from the classroom. This right let families take breaks during school time. He also warns teachers that permitting absence will lead to fines.

"Any time out of school is possible to damage a child's education." Someone at the Department for Education said this weekend. "That is why the government will end the head teachers' right to allow parents taking their children away from classroom."

"This is part of the government's bigger goal to lower the drop-out rate in our schools. There will also be stricter punishment for parents and schools."

A member of the National Association of Head Teachers said this would stop parents from putting pressure on head teachers to permit holidays during school time. "The high cost of holidays during off-school time is still a problem but a child's education is more important than a family holiday," he said.

Appendix F: Reflective Sheet

Writer:

Evaluator:

	Student Identified Errors	Correctly Identified Errors	Overall Errors
Content			
Organization			

THE EFFECTIVENESS OF FORMATIVE ASSESSMENT

Vocabulary			
Grammar			