The University Of Sheffield.

The Impact of Formative Assessment on Young English Learners' Motivation and Achievement in China

By:

Jiawei Zhang

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
Faculty of Social Science
School of Education

March 2018

Abstract

Formative assessment (FA) has been a popular topic of study in education since the 1990s (Black & Wiliam, 1998a; 1998b). In China, the New English Curriculum for primary education in 2011 stressed the importance of FA in language classes (Ministry of Education, 2011). However, empirical studies investigating FA in China are generally lacking (Liu & Xu, 2017). My concern lies in the use and the impact of FA in English language learning in primary schools in China. In that context, this study seeks to identify and scrutinise effective FA strategies that can motivate learners and improve their achievement.

The relevance of FA to English language learning lies in its interactive and dynamic approach, aimed at improving learners' English by offering different strategies for enhancing learning. The FA strategies in my study include sharing learning aims, feedback, questioning, peer-assessment and self-assessment. They were selected from previous works on FA frameworks (Heritage, 2009a, 2009b, 2010; Torrance & Pryor, 2001; Wiliam & Thompson, 2007). I summarised seven motivation theories and three motivation frameworks in language learning (Dörnyei, 1994; Williams & Burden, 1997; Dörnyei & Ottó, 1998) in order to draw up a new framework of motivation in language learning that can be applied in this study.

Mixed methods were used. Data were gathered via quantitative questionnaires for learners exploring motivation, use of FA and attitudes towards FA; quantitative English language quizzes to measure student achievement during the fieldwork; and qualitative interviews with teachers and students regarding their perceptions of FA. Two teachers and 196 11-year-old students from four classes in Jiangsu Province in China took part in the study for two and a half months.

The results showed that FA strategies have valuable application in this context notwithstanding the prevailing exam-oriented education culture in China. FA use was significantly positively correlated with students' motivation and achievement, strikingly so in girls in the lower achievement bracket. There is a need for well differentiated FA due to the varied impact of FA strategies in different student groups. As a lighthouse, FA is not merely an assessment in class, but also a learning habit that guides both teacher and student in teaching and learning. With the findings, this thesis contributes an empirical research study in English of the impact of FA in an underrepresented education sector: primary education in China.

Acknowledgements

I would like to express my special thanks to my supervisor: Dr Tim Herrick. It has been an honour to have a supervisor like him: kind, patient, and supportive. Whenever I have a new idea or get confused, his guidance, advice, and suggestions can always lead me to the next step. I am hugely appreciative of all his contributions of time ideas to this thesis, especially during the last phase of editing and polishing. I could not have completed the project without his enormous help and guidance.

I would like to thank all my participants in the present research because I could not have carried out this study without their vital contributions: the two English teachers in the participating school who offered their time and trust to me as a researcher for an institute outside of China; the 235 students who participated in the research and the 196 in this group who gave me their consent to produce this work with their information. My heartfelt thanks to you all. I would also like to say thank you to Lisa Clarkson who proof-read my thesis thoroughly and provided English language copy-editing help.

I would like to thank my colleagues and friends in the School of Education at the University of Sheffield. I own my thanks to Jiaqi Wei, Yijia Zuo, Wan Angsu-orn, Jaime Chung, Xuexin Guo, and Simin Sasani, and many others who made my PhD years a pleasurable experience and supported me during some tough times in this pursuit.

I would like to thank my friends at the Confucius Institute in Sheffield. I am grateful to have worked with Dr Lucy Zhao, Professor Li Xiao, Professor Feng Ju, and Dr Xinqun Hu during the past five years. In the UK, I have spent most of my spare time with my friends at the Confucius Institute: Fengyong Zhang, Aina Jiang, Qing Xie, Ryan Blackett, Tom Clark. They were always my ‘safe house’ when I was tired of PhD work.

I would like to thank my friends from PETD (People for the Ethical Treatment of Doctors, Social media influencer in China): Lili Sheng, Yuanjun Lu, Han Yang, Aeon Lee, Yan Ma, Haochen Hu, Diya Dou, Liming Jiang, Hui Shao, Di Wang, Zhixian Yu, Jianjia Ma, for the time we travelled together and the songs we sang together.

I would like to thank my best friend Lynn Li, and my parents Yibing Zhang and Chengxia Shu. Words cannot describe how thankful I am to have you in my life.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

B	boy
CET	College English Test
C-Q	close-ended questioning/questions
CF-F	confirmative feedback
CR-F	corrective feedback
CPD	Continuing Professional Development
EFA	Embedding Formative Assessment Project
EFL	English as a Foreign Language
ELFA	English Learner Formative Assessment Project
EM	Extrinsic motivation
FA	Formative Assessment
FAL	FA-lacking environment
FAR	FA-rich environment
G	girl
GCSE	General Certificate of Secondary Education
GPA	Grade point average
HA	Higher-achieving group
ICT	Information and Communication Technologies
IM	Intrinsic motivation
KMOFAP	King's-Medway-Oxfordshire Formative Assessment Project
L2	Second language
LA	Lower-achieving group
LT	teacher who created FA-lacking environment
MA	Medium-achieving group
N-PA	negative peer-assessment
NCEE	National College Entrance Exams
NEC	New English Curriculum
O-Q	open-ended questioning/questions
PEAL	Primary English Assessment for Learning Project
P-PA	positive peer-assessment
RT	teacher who created FA-rich environment
S-F	suggestive feedback
sa	self-assessment
SA	Summative Assessment
SLA	sharing learning aims
TASK	Teacher Assessment at Key Stage Project

*abbreviations are arranged in alphabetical order

Chapter 1 Introduction

1.1 Why begin this study?

Much research in recent years has focused on formative assessment (FA) in language teaching and learning. Wiliam (2006) underscored that formative assessment and summative assessment (SA) are fundamental terms in education. Formative assessment is becoming increasingly important in China as well. In June 2001, a new curriculum system was adopted into the education system in China - the New English Curriculum for Chinese Primary Schools and Junior/Senior Middle Schools (New English Curriculum/NEC). The new system stressed the importance of FA in language learning.

However, in practice, FA was not implemented as well as expected. Consensus from some studies is that there was a mismatch between curriculum intentions and school realities caused by the lack of guidance on how to implement FA successfully (Liu & Xu, 2017; Gu, 2014; Black, 2015). In 2013, when I asked primary school English teachers about how they use FA in their teaching, they could not satisfactorily explain what FA is. That response drove me to conduct a study for my master's dissertation on the utility of FA in China. This study found that sharing learning aims, questioning, feedback, peer-assessment and self-assessment are FA strategies which are often used in language classes in China. The findings of my master's dissertation explained the use of FA and its impact from teachers' perspectives; however, it was apparent that further investigation was needed. FA's 'impact' on students to improve language learning was too general a relationship. What I wanted to investigate in the present study was the specific relationships between FA strategies and learners' motivation and achievement, using authentic data collected in a naturalistic classroom setting.

Being a primary school student in China is different from the pupil experience in the UK. Therefore, the findings in studies conducted in the UK context may not be readily applied in China. Due to population pressures, the typical classroom is normally packed with pupils, but it is well organised. FA strategies that are used in a class with 30 or fewer students may differ from those in a class with 50 students or more. The school day starts at 7:00 in the morning and finishes around 5:00 in the afternoon. There is a one- or two-hour lunch break depending on the season and school rules. It is very easy for students to get sleepy or tired if they sit in the classroom all day, over eight 50-minute lessons. The students may need more stimulation beyond the normal teaching methods to promote their learning. Moreover, children aged from 10 to 15 develop and grow more rapidly than at any other developmental stage, especially cognitively (Barratt, 1998). This is the age range when most children in China start to learn English as foreign language. Therefore, 196 10/11-year-old children from four classes in Jiangsu Province in China took part in the study for two and half months. In this period of rapid development, children are more likely to improve if they experience the most appropriate teaching and assessment strategies. Hence, this study pursued empirical evidence of the impact of an alternative to the traditional teaching, namely, formative assessment's influence on motivation, in the context of primary schools in China.

1.2 My positioning

Being a university undergraduate who majored in teaching English to young learners, the training and education about how to assess students was merely paper talk before entering the real classroom. In my MA study, I majored in in Applied Linguistics with TESOL, and classroom assessment became a more important issue due to teaching part-time at a language institute. When attempting to apply the theories into practice, the difficulties emerged naturally.

First of all, it was difficult to take classroom assessment into consideration as an inexperienced teacher because the priority at this point was knowledge delivery. To ensure the completion of the teaching, there are many other things need to be done before assessing students, such as classroom management and lesson preparation. It will take plenty of time for a new teacher to 'settle down' in a classroom. Therefore, most of the assessment will be given out of instinct rather than well prepared. The assessment theories in books and lectures were not easy to digest and there was no time-saving straightforward toolkit to explain what to do in the class. Secondly, the impact of certain feedback techniques is worth further investigation. In my teaching, many students, especially the top students, who received a lot of positive feedback from teachers and peers, did not show improvement in their achievement. On the other hand, criticism appeared to play a positive role in promoting students' learning. This posed questions in my mind: can confirmative feedback or other strategies be beneficial all the time, and how much confirmative feedback will be enough? Therefore, the dilemma emerged in my own teaching which triggered the exploration of the use of FA, and the connection between FA, motivation and achievement. The change in my professional identity from front-line teacher to academic researcher made this exploration possible.

1.3 What has been studied

The aforementioned thinking was the impetus for the research. A grounding in the literature was needed to support the feasibility of doing the research as a PhD project.

Formative assessment was again brought to public attention in the late 1990s by Black and Wiliam (1998a; 1998b), two decades after Scriven (1967) coined the term and Bloom (1968) applied it to teaching and learning. Although FA does not have a universally accepted meaning, many researchers regard it as a classroom activity that improves students' learning via a process of collecting data on students' performance and then adjusting the teaching accordingly.

Studies of FA that are reviewed in this thesis include research into FA frameworks (such as Heritage, 2009a; 2009b; 2010; Torrance & Pryor, 2001; Wiliam & Thompason, 2007), FA and motivation (such as Cauley & McMillan, 2010; Harris, 2007), and FA and achievement (such as Andersson & Palm, 2017; Black & Wiliam, 2004; Wiliam, Lee, Harrison, & Black, 2004). For FA frameworks, different models have been developed, yet several common strategies were identified and these are sharing learning aims, questioning, feedback, peer-assessment and self-assessment. Therefore, these became the core strategies studied in this thesis. Regarding FA and motivation, research has found that they are closely related, however

most of these studies were conducted in Anglophone environments. Studies have also explored the relationship between FA and achievement, with varied findings. Positive impact of FA on students' achievement was found in some studies, such as in Black and Wiliam's (2004) and in Andersson and Palm's (2017). However, some other studies found that FA had little or no positive impact on achievment, for instance, in Heitink *et al*.'s research (2016) and in Lam's (2016). Therefore, the impact of FA on achievement remained a subject of debate and has been questioned by other researchers (Dunn & Mulvenon, 2009; Bennett, 2011). In FA research, some studies have investigated particular FA strategies, such as questioning (Jiang, 2014); feedback (Harris, Brown, & Harnett, 2014; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Lee, Mak, & Burns, 2016); peer-assessment (Cheng & Warren, 2005; Wang, 2014); and self-assessment (Butler & Lee, 2010; Butler, 2018). It is clear that FA cannot be represented by a single strategy, and it is not an 'empty' concept that has no established, concrete activities. Therefore, this study explores the strategies and approaches mentioned above, with a particular interest in how they could contribute to an understanding of FA used for English language learning in primary schools in China.

Chinese scholars started to pay attention to FA due to the great influence of studies conducted in the UK (such as the the research by pioneers Paul Black and Dylan Wiliam *etc.*), in the USA (such as some recent work by Dylan Wiliam *etc.*), in Europe (Jonsson, Lundahl, & Holmgren, 2015; Peng, 2017 *etc.*), and in Hong Kong (Carless, 2012; Carless & Lam, 2014; Lam, 2016 *etc.*). The updated version of the NEC (Ministry of Education, 2011) also promoted FA as the new assessment method in teaching and learning. Luo published a pioneering book about FA in education in mainland China (Luo, Huang, & Xiaolei, 2014). However, there is a mismatch between curriculum intentions and school realities, as well as the challenge of teachers' pressures within an exam-oriented culture (Liu & Xu, 2017; Gu, 2014).

The characteristics of teaching and learning in China include: an exam-oriented culture, a conception of learning in which effort is central, hierarchal teacher-student relationships, collectivism values, big class sizes, and an eagerness for quick successes and instant benefits *etc.* (Chen, 2015). Some of these characteristics conflict with the nature of FA. For example, FA's enduring feature is that it does not involve simple 'one-off' administration. FA activities are time-consuming and require a lot of effort from teachers. Moreover, big class sizes in China have increased the workload on teachers, exacerbating the problem of the time and effort needed for effective FA (Hu, 2002). Also, English classes in China are often 'quieter' because the students are not expected to speak up (Carless, 2012), so FA via oral interaction cannot be established easily. Therefore, the potential for positive impacts of FA in Chinese schools needs more exploration and evidence.

1.4 Where the research gap is

Although FA has been a popular topic in educational reform in China, the literature about FA in primary education is generally lacking, especially empirical studies (Liu & Xu, 2017; Chen, 2015; Carless, 2012; Huang & Luo, 2014). There have been ample studies about FA conducted in higher education settings (*e.g.*: Chen *et al.* (2013; 2014); Xiao & Carless's (2013)), as well as secondary schools (*e.g.*: Gu (2014); Dello-Iacovo (2009)).There have been far fewer studies in primary education. In particular, the use of day-to-day oral feedback is

relatively underexplored (Xu & Carless, 2016). The positive impact of FA on learning has been asserted by many researchers, such as one of the most influential articles by Black and Wiliam (1998b). However, the positive relationship has been questioned by other studies which found little or no positive impact, or which reviewed the evidence in previous work and found it lacking (Dunn & Mulvenon, 2009; Bennett, 2011). Hence, this study will provide further empirical evidence about the impact of FA which can contribute to the existing debates. Moreover, many studies have reported the impact of FA strategies on motivation, yet did not connect the benefits with learning achievement (Liu & Xu, 2017). Researchers often treated FA as a whole, neglecting to differentiate the effects of each strategy within it; or treating one particular strategy as the representative of FA. Almost every FA strategy has been investigated separately, but not many studies have not many researchers have considered a range of FA strategies in depth in one study. Many studies about FA in China merely focused on 'empty' theoretical discussions. The positive impact of FA was stressed repeatedly, yet there was little empirical evidence for it, especially in the literature published in Chinese. Therefore these studies are not often cited or relied on in this study. In relation to the Hong Kong system, Brow *et al*. (2009) concluded that there was a significant barrier for teachers to change the purpose of assessment from the traditional 'assessment *of* learning' to 'assessment *for* learning'. Lastly, in Anglophone literature, studies of FA were mainly conducted in Anglosphere systems. This thesis contributes a case study for other researchers who may want to compare FA in different cultures.

The aim of this study is to investigate the impact of FA on learners' motivation and achievement in the language learning process in the context of Chinese primary schools. It would show how FA, originally a 'western concept', may interact with Chinese school culture; meanwhile, it is an opportunity for researchers who cannot read Chinese to understand how FA fits into other contexts and makes a difference. Three research questions were raised to explore the quantitative relationships between the three themes of this study:

1) Is there a relationship between FA and students' motivation?
2) Is there a relationship between FA and students' achievement?
3) Is there a relationship between students' motivation and achievement?

Two qualitative research questions were also raised because the numerical evidence alone cannot answer the questions regarding 'why' relationships exist.

4) What is the impact of FA from English students' perspectives?
5) What is the impact of FA from English teachers' perspectives?

The data were collected from English classes in primary schools in China, and written up in English as a PhD thesis at the University of Sheffield. This thesis used authentic recourse to explore the impact of FA on students' motivation and achievement in China, which fills the research gap identified earlier. The first three research questions ascertained whether there was a relationship between FA, motivation and achievement, and the following two questions investigated what teachers and students think of the relationships. According to the characteristics of the students, the participants were divided into different achievement groups, and they experienced different FA environments. Gender was another variable that

was explored. Therefore, each research question was elaborated in detailed sub-questions (see 3.2.1). With these five research questions, the impact of FA on students' motivation and achievement can be explained using numerical evidence as well as the subjective perceptions gathered from the participants.

1.5 How this thesis is structured

A traditional thesis structure has been chosen. This section gives an overview of the study and signposts the content and objectives of the seven chapters :

Chapter 1: Introduction

This chapter starts by describing the researcher's initial intentions for implementing this study (1.1). My positioning (1.2) introduces the background of the researcher, which add more information to the reason for conducting this study. Four years is not a short period of time, so it is challenging to devote oneself to one single subject unless there is a strong interest in it. Academic research differs from other study and writing. It requires strong grounding evidence to support the rationale for doing the research. Therefore, exploring the existing literature on FA was a priority (1.3). (This scoping exercise was different from the literature review described in Chapter 2 where relevant previous published works were systematically searched for, identified and reviewed.) A map of the literature was then drawn using the introductory information and highlights the research gap described in the following section (1.4). It is essential in a doctoral thesis, which is an original piece of research, to identify areas which have not been studied or not been studied adequately. The last item in the introduction is the map of the thesis in this section (1.5). It should be easy for readers to follow the structure of a thesis with such a traditional pattern. This section also explains the reasons behind the structure and reveals the connections between chapters. Using a series of questions, Chapter 1 outlines the 'footprints' of the whole thesis and guides readers into the main body.

Chapter 2: Literature review

The main body of the thesis starts from the literature review. This chapter gives a comprehensive review on three main themes: formative assessment, motivation and achievement. Firstly, for each theme, the concepts are identified and defined (2.1.1, 2.2.1, 2.3.1) because it is important to know exactly what is being studied.

"If I have seen further it is by standing on the shoulders of Giants". The FA framework in this study is developed and confirmed after reviewing previous models (2.1.2). Due to the scope of this study, literature on FA was specified in the context of China (2.1.3) and in the field of language learning (2.1.4). It was found that although FA in language learning was studied by many researchers, in China it still needed more exploration. Teacher's viewpoints on FA are considered and presented because they are also important participants in this study (2.1.5). Their contributions also provide valuable material for the later discussion of teachers' perceptions of formative assessment. Considering the identified research gap in FA, the subsequent section focuses on the existing literature on FA in primary education (2.1.6). The current study did not select the three research themes randomly. The literature about the

connections between FA and motivation/achievement highlight the ongoing debate (2.1.7). The review reminds the reader that this study can make a contribution to this literature. The next section asserts my personal expectations about the use of FA with technology in the future (2.1.8). It argues that it is possible and efficient to use FA with technology, but only if local conditions allow. In many classes in China, the equipment to apply computer-assisted FA is provided. It is other issues, such as teachers' perceptions and the special cultural context, that have hindered the implementation of FA in China. Therefore, the delineation of the obstacles and challenges (2.1.9) enhanced one of the aims of this study – to provide practical and 'localised' suggestions to teachers.

In the introduction to studies on motivation, the main theories are elaborated (2.2.2). They are Expectancy-value Theory, Attribution Theory, Self-efficacy Theory, Achievement Theory, Goal Orientation Theory, Goal Setting Theory and Self-determination Theory. On the basis of these theories, a synthesis motivation mechanism, with elements of action, individual and environment, is summarised. Three motivation frameworks in language learning by Dörnyei, Williams & Burden, and Dörnyei & Ottó are compared (2.2.3). The elements of the theories are compared against the frameworks and then selected and rearranged into a new framework tailored for this study (2.2.4). The new framework provides the foundation for the questionnaire about motivation which was applied in this study.

Turning to the notion of achievement, the distinction between achievement, attainment and progress is described, prompted by the thesis confirmation review (2.3.2). It was true that many people use achievement and attainment interchangeably, but they can mean slightly different things. Clearer distinctions enabled clearer exploration of the topic. As with the literature on FA, the literature on achievement was specified in the context of China (2.3.3) and in the field of language learning (2.3.4). On one hand, it suggested corresponding information in a parallel structure; on the other hand, the achievement in China and in language learning appears to be different from other contexts. When the connections between FA and motivation/achievement were introduced, the relationship between motivation and achievement was not mentioned. Therefore, a section on this relationship (2.3.5) completes the triangular relationship among the three themes and, more importantly, provides information for the discussion in Chapter 6. Given that studies have found significant gender differences (in achievement and in perceptions of FA), and gender is a variable in the current study, the literature on gender difference in achievement is thoroughly reviewed (2.3.6). The last section compares the different ways of measuring achievement (2.3.7). Combining useful information from each method, a suitable achievement measurement is suggested for this research.

The review for each research theme ends with a summary of the literature in each section (2.1.10, 2.2.5, 2.3.8).

Chapter 3: Methodology and Methods

This chapter describes the issues concerning the methodology, including the research aims and questions; philosophical positions; the choice of research methods; the development of research instruments; the background of the participants; the procedure of data collection;

the research ethical issues and last but not the least, the limitations of the methodology and corresponding reflections.

Methodology and methods is in many respects the most important chapter as it builds the foundation for the whole thesis, and determines the validity and reliability of the research. Interestingly, the chapter begins with sharing research aims (3.2.1), mirroring one of the FA strategies in this study. Similar to an interview participant's observations about sharing learning aims, the research aims provide directions and guidelines for the whole research. The subsequent chapters about data analysis and discussion both use research questions as the chapter structure. In other words, research questions are the core of this study. Because several of the concepts in this research, such as motivation and achievement, may mean different things to different people, it was important to produce operational definitions to 'visualise' them as doable and measurable concepts which can be applied in this particular study (3.2.2).

One special aspect of doing research in the UK is the statement about personal ontology and epistemology, which is usually neglected when doing research in China. Therefore, in the exploration of philosophical positions (3.3.1), the thinking about the entity and the ways of seeing the entity are clarified and deepened. The philosophical position also helps to partly explain the choice of methods (3.3.2) and the discussion in Chapter 6 because as the traditional Chinese saying says: "There are a thousand Hamlets in a thousand people's eyes". The discussion can only represent the viewpoint of the researcher with her own ontology and epistemology, and the research methods are a path to the final discussion. On the basis of the research methods and the information obtained from the literature review, the research fieldwork instruments were developed and this process is described in 3.4.

As to the actual fieldwork, a description of the background (3.5), including participants and their community, gives a brief sense of the research environment. There is an explanation of the sampling methods which sought to ensure the validity and reliability of the findings. The section about procedure introduces the steps taken in the fieldwork (3.6) with a detailed research timetable. The next section (3.7) explains the how each data set was analysed. Each step is explained in detail and 're-playing' the fieldwork in this way was helpful to reflections on the methodology (3.9). The ethical considerations are explained and discussed (3.8). The context of this study is cross-cultural research in primary education, so attention was paid to the cultural differences of doing research in China. Due to the nature of the participants, there is a section about the ethics of doing research with children. Other ethical issues, including anonymity and confidentiality, are also discussed.

Chapter 4 & 5: Quantitative and Qualitative Data Analysis

These two chapters were originally intended for a single chapter on 'Data Analysis'. However, it was challenging to present and analyse these rich datasets together. Data analysis is therefore organised in the order of research questions: Chapter 4 presents quantitative data analysis which answers research questions 1, 2 and 3; Chapter 5 presents qualitative data analysis which answers research questions 4 and 5.

Chapter 4 starts with details about the participants (4.2). By introducing the composition of the group of participants, this section seeks to show that the participants were equally distributed across the conditions. Then the basic information about each theme, namely achievement (4.3), motivation (4.4) and FA (4.5) are introduced. The section about achievement presents the dynamic changes over the 10-week fieldwork period. It also explains the categorisation of the participants into three achievement groups according to their academic performance during the fieldwork. The next section describes the validity and reliability of the questionnaire about motivation and it labels the factors in motivation by SPSS which are used in the discussion chapter. Apart from these unique subject-related analyses, the rest of each section shows the quantitative data in a similar structure: overall data in the pre-test and the post-test; data in different settings: FA environments, genders and achievement groups; and the variation between the pre-test and the post-test in the aforementioned settings. Only where quantitative data are arranged thoroughly in such a homologous and structured pattern, can the comparison between the themes be comprehensive and organised. In the subsequent sections, the relationship between FA and motivation (4.6), between FA and achievement (4.7), and between motivation and achievement (4.8) are presented in terms of achievement group, FA environment, gender, and crosstab of achievement group and gender. The crosstab was newly added due to the results of analysis which found that girls in the lower achievement group had the most significant relationship with other variables in every comparison.

The structure in qualitative data analysis is simpler to that in quantitative analysis. It also follows the research questions, hence there were two sections: students' (5.2) and teachers' (5.3) perceptions of FA. In each section, five FA strategies are discussed in detail since teachers and students gave their opinions at length and there are many comments worth quoting. Finally, each chapter ends with a summary (4.9, 5.4).

Chapter 6: Discussion

The discussion chapter links the synthetic data from the two findings chapters in this study with previous works. The discussion of findings is arranged under the main themes of FA strategies (6.2): sharing learning aims, questioning, feedback, peer-assessment, and self-assessment; motivation and achievement (6.3). The relationship between motivation and achievement is discussed in the different settings, namely achievement group, FA environment, gender and crosstab of achievement group and gender. On the top of this structure, there is a few sections that discuss about lower-achieving girls (6.4) due to the unexpected and significant findings in this group; the unexpected information of the other impact on motivation and achievement (6.5); particular points for consideration in relation to the findings (6.6). These were unforeseen and important data and they provide new knowledge by either deepening the findings or expanding the margin of the work, because it was beyond the original framework of this research. In the current study, data is not reported selectively, to avoid bias leading to untrustworthy or incomplete findings.

Chapter 7: Conclusion

Similarly to Chapter 1, the concluding chapter answers five questions. First, it describes the impetus that pushed me through four and a half years to finish the project. The desire and

intention were to show the significance of this study, which meets the needs of teachers, students and researchers (7.1). Secondly, a summary narrative about the research project was made from the beginning to the very end (7.2). This was different from the research procedure introduced in 3.6 which only focused on the fieldwork. As a researcher, I shared the experience and beliefs of conducting research with other readers to provide a reference for their future works. Hopefully, this will encourage more work to be done in this field to improve teaching and learning in China, because there are many other researchers in similar situations, for example, doing research about FA, or conducting fieldwork in China. This is a possible contribution that I can make as a researcher, as well as a Chinese. The next section (7.3) reminders reader of the five research questions by answering them with solid evidence from the previous discussion and a succinct response in the end. The subsequent section presents the most critical part of this study - the original contributions (7.4). One of the reasons that I started this project was to offer practical suggestions to teachers based on the findings. Hence, the suggestions about using FA, especially regarding using different FA strategies with different student groups, are raised in 7.5.1. In addition to advice for front-line teachers, some suggestions are made for teacher training (7.5.2). In many research studies, recommendations based on findings are often given to policy-makers. However, for this study, due to certain factors such as sample size and complex issues in the Chinese education system, suggestions remain at the micro level of environment changes (Carless, 2012), which include internal and external school-based support and so on. The section about limitations (7.6.1) examines issues caused by practical constraints or a lack of research experience and thoughtful considerations; the section about reflection (7.6.2) emphasises the personal thinking about the topic and the implications of this study. This included thinking about what could be improved if there is a second chance to conduct the same research. If there is an opportunity to extend this study on the basis of this research experience and findings, several directions are suggested (7.7).

The last chapter uses a similar format to the first chapter. It echoes the beginning in content as well as format. It addresses such questions as - why begin and complete this study, what has been studied, and where to next? More importantly, it gives a sense of having completed the 'circle' of the whole thesis.

There are several appendices after the main body of the thesis. First, documents and instruments used in the fieldwork are appended along with their translation if the original was in Chinese, such as the questionnaire, the English quiz sample, interview guide for teachers and students and files about ethical issues *etc.* Secondly, the appendices include tables of essential quantitative data that readers can use as a reference when reading Chapter 4 Quantitative Data Analysis.

Having presented the background information as well as an outline reporting structure, the thesis will proceed with a literature review of formative assessment, motivation and achievement.

Chapter 2 Literature review

2.1 Formative Assessment

2.1.1 Defining FA

In research, the definition of the key constructs being studied is crucial both to the research process itself and its practical application. It is difficult to document effectiveness meaningfully if definitions are not clear; therefore, when applying the research it is even harder to translate it to a specific context (Black & Wiliam, 2004; Bennett, 2011). In this section, the definition of FA and related relevant terms will be discussed and compared, in order to build the stepping stones for the whole thesis.

Assessment vs. Measurement/Evaluation

Broadly speaking, the terms assessment, measurement and evaluation are used interchangeably by some practitioners. They seldom pay attention to the differences and treat all of them as forms of 'tests', especially in teaching practices. However, the research suggests that they do not have the same meaning.

According to Clapham (2000), assessment is used either as an umbrella term for all testing and assessment methods, or a term that refers to informal methods that differentiate testing and alternative assessment. Assessment also concerns the 'environment' in which the individual has developed, because the learner's interaction with the environment is seen as important to the learning process (Bloom, 1968). It examines information about both the individuals and the environment to understand a phenomenon, and then uses the data as a basis to predict outcomes or control outcomes. Testing refers to formal or standardised tests, which are specially developed forms or instruments of assessment to gain quantitative data from various sources.

This is similar to what measurement does in education: to assign a quantitative value to a quality. For example, the quantitative score of a child is used to determine their performance. The environment is generally ignored when in the practice of making measurements, even though the environment is acknowledged in theories of measurement, such as Bloom (1968). Measurement is quantifying data using existing standards or units, showing the facts, but does not include evaluation.

Evaluation, with 'value' as the core notion, makes judgements about the worth or quality of a performance by comparing the outcome to the defined criteria (Ogunniyi, 1984; Huitt, Hummel, & Kaeck, 2001; Kizlik, 2017). Evaluation follows the objectives of the instruction (*i.e.* in teaching or training). Therefore, it differs from individual to individual and school to school. The environment, seen as a potential source of error in measurement, is assumed to be a major source of the changes (Bloom, 1968). Evaluation emphasises the classification of individuals in order to grade them, grant certificates, or promote them in the workplace *etc*. It is a useful tool to maintain quality control in education. To encapsulate the distinctions in single words, as shown in Table 2.1, we collect information via assessing learning; we

quantify the information via measuring distance; and we make judgements via evaluating results.

Table 2.1 Comparison between achievement measurement and evaluation

Dimension	Assessment	Evaluation	Measurement
Purpose	Diagnosis	Judgement	Measurement
Timing	Continuous	Summative	Discrete

Assessment is not about numbers; it concerns feedback. Broadly speaking, assessment refers to all the activities undertaken by teachers, or by students, which provide information to be used as feedback to "modify the teaching and learning activities in which they are engaged" (Black & Wiliam, 1998b, p 2).

Formative assessment vs. Summative assessment

Scriven (1967) firstly proposed the terms formative assessment (FA) and summative assessment (SA), to explain the evaluation in curriculum; then Bloom *et al*. (1968; Bloom, Hastings, & Madaus, 1971) subsequently applied these two terms to student learning. FA is used to provide feedback and correctives from small learning units in the teaching-learning process (Bloom, 1969) to improve what the student, teacher and curriculum designer do (1971). Since then, FA and SA have become fundamental concepts in education (Wiliam, 2006).

To distinguish SA from FA, one obvious characteristic of SA is measurement for accountability purposes. It is more likely to be a measurement (as specified in Table 2.1) than an assessment. It usually appears at the end of a learning period with a graded form to judge the outcomes from the learning that has occurred (Torrance & Pryor, 1998; McAlpine, 2002). Stiggins and Chappius (2006) refer to SA as 'assessment of learning' and FA as 'assessment for learning'. However, the boundary between FA and SA is not always clear-cut. Sometimes, SA can be used formatively - for instance, using the data from SA to give feedback to improve learners' achievement (Carless, 2012). FA can be presented in a summative format, such as a group competition in the class with scores. Black *et al*. (2003) suggest that it is difficult to use FA and SA separately. According to Rowntree (1977), identifying the *intentions* of assessment, rather than the form of assessment, is a good way to categorise assessment as formative or summative.

The importance of SA, or assessment for accountability, is undeniable. Gordon *et al*. (2014) advocate a balance between FA and SA. Each serves a different but complementary function in education. As Rea-Dickins and Gardner (2000) argue, the distinctions between FA and SA are actually not as straightforward as they seem. The interplay between FA and SA is highly complex.

'Formative' in classroom assessment probably means different things to different people. There are several core ideas in FA which appear in most of the FA definitions. First of all, the FA is an activity between teacher and students. The assessment is administered by the teacher and used on the students, and/or administered by the students and used on themselves or their peers. It falls within the school teachers' responsibility but can also be undertaken by their students (Black & Wiliam, 1998a). It can be characterised as a social

interaction or instruction between teacher and students (Torrance & Pryor, 1998; McManus, 2008). Rea-Dickins (2001) suggests that both the teacher and student are the key stakeholders in the assessment process.

Secondly, collecting information about students' performance is required. The continuous flow of information (Stiggins, 2002; Heritage, Kim, Vendlinski, & Herman, 2009), is gathered about aspects of learner performance (Rea-Dickins, 2001) and can be used as evidence later (Black, Harrison, Lee, Marshall, & Wiliam, 2003). Instruction is the tool used most often to gather the information about performance and to identify gaps and diagnose students' strengths and weaknesses (Airasian, 2001; Wininger & Norman, 2005; McManus, 2008).

Thirdly, as Wiliam has observed (2006), FA information is used to make changes, to modify teaching. Modifications have a range of purposes: to make the teaching more appropriate for students (Bachman & Palmer, 1996); to meet learning needs (Black, Harrison, Lee, Marshall, & Wiliam, 2003); to improve student performance (Wininger & Norman, 2005; McManus, 2008; Airasian, 2001); and to advance rather than check on student learning (Stiggins, 2002). Black (1995) concludes that FA information is used to modify the work to improve the effectiveness of teaching and learning, rather than merely verify the results.

The last idea is that FA is a continuous process, which Huinker & Freckmann (2009, p 7) refers to as "ongoing and integral process". Heritage *et al*. (2009, p 24) echoes this, describing FA as "a systematic process to continuously gather evidence and provide feedback". Therefore, FA cannot be seen as one-off activity, but rather as a continuing process that requires the efforts of both sets of key stakeholders.

Among all the definitions, there are two that are best applied in this study:

> *An assessment activity can help learning if it provides information to be used as feedback by teachers, and by their pupils in assessing themselves and each other, to modify the teaching and learning activities in which they are engaged. Such assessment becomes formative assessment when the evidence is used to adapt teaching work to meet learning needs* (Black, Harrison, Lee, Marshall, & Wiliam, 2003, p 2).

> *An assessment functions formatively to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers to make decisions about the next steps in instruction that are likely to be better or better founded, than the decisions they would have made in the absence of evidence* (Wiliam, 2011, p 42).

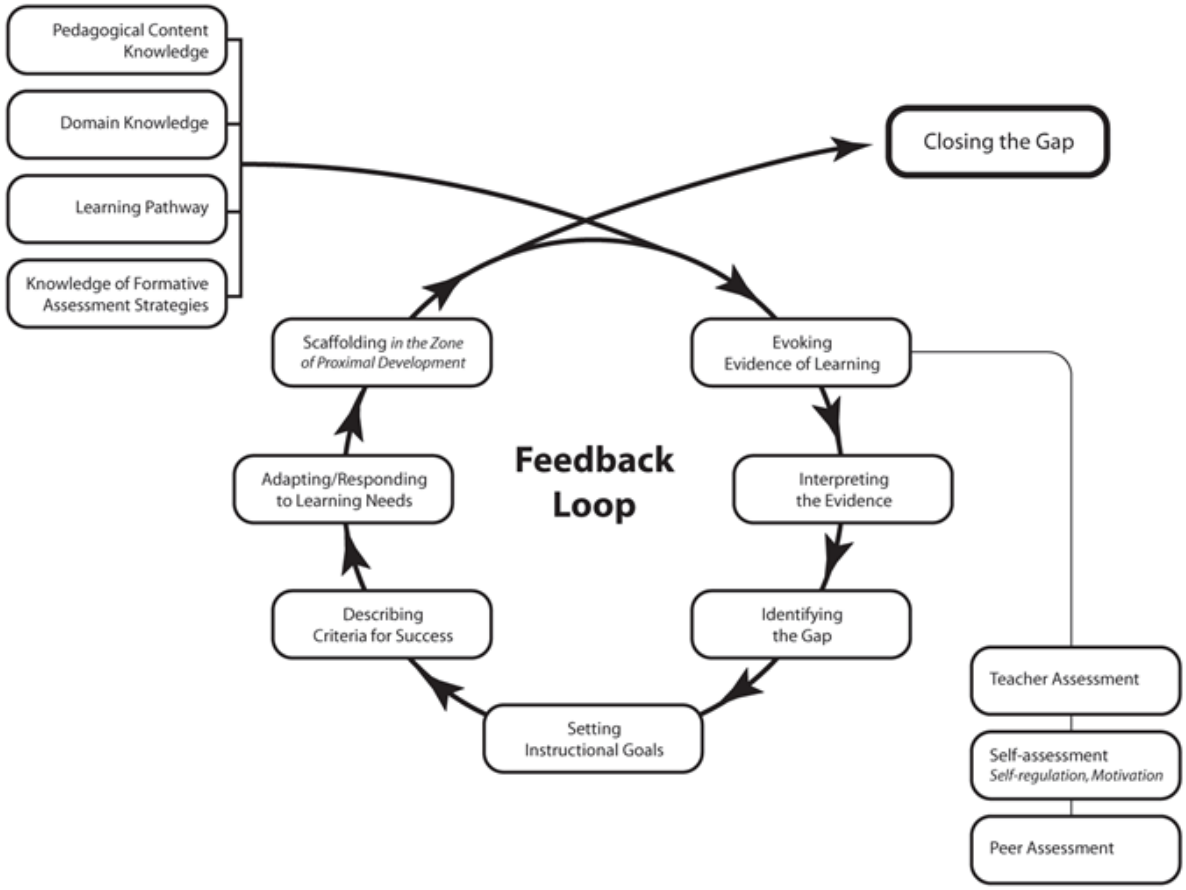
To sum up, FA is an ongoing and dynamic activity that collects information about students' performance and the progress of teachers and students, via the use of instruction, observation, or conversation *etc.* The information gathered is used to identify gaps between performance and learning needs, and then make changes in the subsequent teaching and learning process to improve students' learning thereafter. However, it is worth mentioning that FA in this study mainly focuses on assessment within the classroom. Other types of FA, such as portfolio assessment, are not investigated here.

2.1.2 FA frameworks

Several FA frameworks have been developed to help practitioners implement better FA in teaching and learning. These frameworks have a few elements in common, such as collecting information, giving feedback, and adjusting the teaching. Only the details specified for the process are different.

Heritage (2009a; 2009b; 2010) proposes an FA model which focuses on elements of FA: "learning progressions, including learning goals and success criteria, identifying the gap, eliciting evidence of learning, teacher assessment, teacher feedback, and student involvement". As shown in Figure 2.1, the loop can be visualised with a logical sequence of features in FA, as follows: with the pre-acquired pedagogical content knowledge and FA knowledge etc., identify the learning gap with evidence that may collected by teacher assessment and student-led assessment; on the basis of the gap, set learning goals and success criteria; finally, close the gap by adapting the learning needs, or set new learning goals if the gap is not closed. However, some of these elements overlap. For instance, setting instructional goals is commonly connected to describing criteria for success.

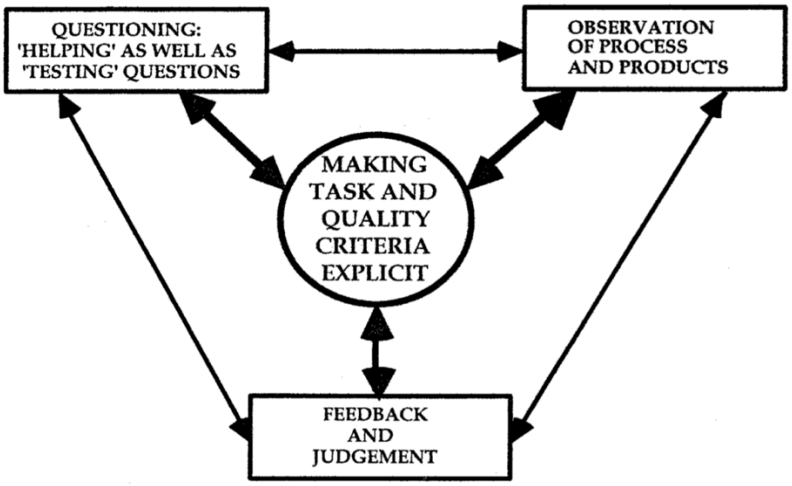
Figure 2.1 FA Framework by Heritage (2010, p11)



Torrance and Pryor (1998) developed a descriptive and analytic framework of the processes of FA in the classroom (Appendix 1). However, this process is too specific to apply in a wide range of classes. In 2001, Torrance and Pryor modified the framework after conducting the *Investigating and Developing Formative Teacher Assessment in Primary Schools (Primary*

Response Project), which was designed to investigate and develop FA in primary schools. They developed an FA framework in practice, which is illustrated in Figure 2.2 below.

Figure 2.2 FA Framework by Torrance & Pryor (2001)



They sort the concrete examples in Appendix 1 into four main categories, shown in Figure 2.2. Similar to Heritage's FA loop, Torrance and Pryor emphasise criteria of learning goals and teachers' feedback. Meanwhile, they highlight classroom observation in the process of using FA. The teacher observes the students' learning process and learning outputs. However, the observation and questioning happen at the same time. Therefore, in the FA framework used in this thesis, questioning - action which can draw the learner's attention - will be retained. Observation will be treated as a part of the feedback loop in Figure 2.1, because this is an enduring process which happens along with questioning, feedback and other FA strategies. This framework emphasises the perceptions of teachers, but does not pay attention to learners' assessment. In other FA frameworks (Figure 2.1;Table 2.2), peer-assessment and self-assessment play significant roles.

Four areas are focused on in Black and Wiliam's work (1998b): questioning, feedback, sharing criteria, self-assessment; later, Black *et al*. (2002; 2003) explore four strategies in FA: questioning, feedback through marking, peer-assessment and self-assessment by students, and the formative use of summative tests. Drawing together Wiliam's previous work (2000; 2007a; 2007b), Black and Wiliam (2009) suggest five main types of activity: sharing success criteria with learners, classroom questioning, comment-only marking, peer-assessment and self-assessment, and formative use of summative tests, and those activities are embedded in the following framework (Table 2.2) developed by William and Thompson (2007).

Table 2.2 Aspects of formative assessment by Wiliam & Thompson (2007)

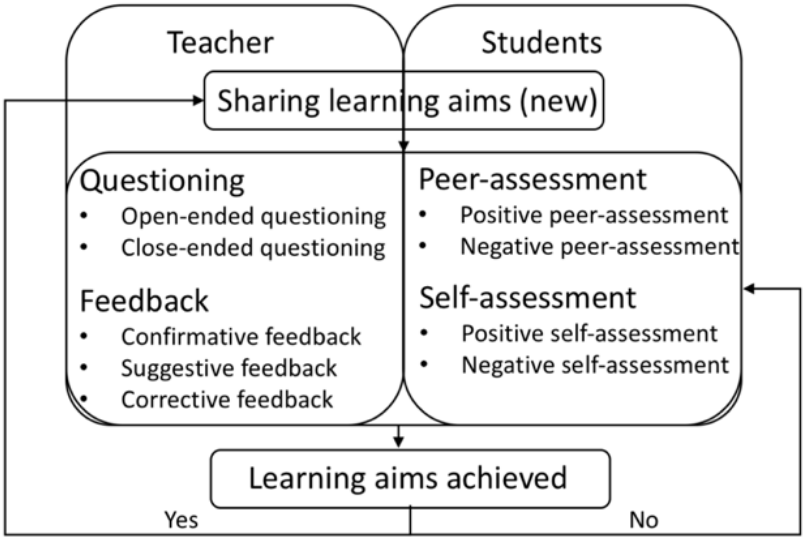
	Where the learner is going	Where the learner is right now	How to get there
Teacher	Clarifying learning intentions and criteria for success*	Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding*	Providing feedback that moves learners forward*
Peer	Understanding and sharing learning intentions and criteria for success	Activating students as instructional resources for one another*	

Learner	Understanding learning intentions and criteria for success	Activating students as the owners of their own learning*

* indicates five key strategies in FA

This framework is a comprehensive model which captures the learning process that Ramaprasad (1983) defined: where the learner is going, where the learner is right now, and how to get there; as well as the agents of FA: teacher, peer and learner. However, the framework used in the present study has been simplified to make it easier to apply in the classroom. It is depicted in Figure 2.3 below. It is grounded in the Wiliam & Thompson's framework and draws on knowledge of English teaching in Chinese primary schools. It can be seen as a handy toolkit for teachers because FA is no longer an abstract concept but five clear strategies with concrete assessing means. The flow in the framework also suggests the assessment timing for the teacher. However, the framework is only a toolkit, the teacher still initiates the classroom instructions.

Figure 2.3 FA framework in the present study



As illustrated in Figure 2.3 FA framework in the present study has strong theoretical support in the main FA strategies proposed by Black, Wiliam *et al*. (2009); Heritage (2009a; 2009b; 2010); and Torrance Pryor (1998; 2001). The flow of assessment starts with setting new learning aims, and it ends with closing the gap and setting further new learning aims. This idea was borrowed from Heritage's framework in Figure 2.1. Setting goals or clarifying criteria, teachers' questioning, and feedback are shared elements in these three frameworks. Both Torrance Pryor (1998) and Wiliam & Thompson (2007) suggest that assessment by the learners is important. Therefore, the main FA strategies in this study are of five kinds: sharing learning aims, questioning, feedback, peer-assessment, and self-assessment.

Following the fieldwork during my MA dissertation, I developed typologies in different assessment strategies. Within the framework of FA, each strategy has its own sub-types. There is some support from other frameworks too. For example, in Torrance and Pryor's work, questioning includes helping as well as testing, which matches open-ended and close-ended questioning respectively. The following Table 2.3 shows the four types of feedback

identified by Tunstall and Gipps (1996a; 1996b). In their system, there are two main classes of feedback: evaluation and description. Types A and B in evaluative feedback focus on rewards/sanction and level of approval respectively. Types C and D in descriptive feedback concentrate on products and processes of learning. However, such a detailed typology of feedback may be too complex to put into practice. In classroom teaching, forms of feedback are always embedded in each other, such as rewarding and specifying students' attainment. It is the *intention* of the feedback that differentiates it. Therefore, in this study, the feedback is categorised three ways - confirmative, suggestive, and corrective.

Table 2.3 Typology of teacher feedback

Evaluative Feedback	Type A	Type B
Positive Negative	Rewarding Punishing	Approving Disapproving
Descriptive Feedback	Type C	Type D
Achievement focused Improvement focused	Specifying attainment Specifying improvement	Constructing achievement Constructing the way forward

Adapted from Tunstall & Gipps (1996a)

To sum up, what Figure 2.3 depicts is a snapshot of a teaching unit with FA. It starts with the teacher sharing learning aims with students. This is a central and crucial element in the FA Framework proposed by Torrance and Pryor (Figure 2.2), and a starting point in Wiliam & Thompson's model (Table 2.2). However, due to the need to promote student autonomy, the responsibility for learning aims is not placed on teachers only. Teachers share learning intentions with students, but do not know where the learning will end up (Black & Wiliam, 2009). The strategies of questioning, feedback, peer-assessment and self-assessment are in a 'pot' and can be drawn on at any time during the learning period. Questioning uses questions to explore deep understanding, at the same time, more time is given to answer. Feedback focuses on how to improve, rather than on how well the students have done. Peer-assessment and self-assessment give students rubrics and time to assess their own and each other's work, identifying priorities for review.

2.1.3 FA in China

At a macro-level in education, China is the first country that developed an examination system (Black, 1998; Lewin & Wang, 2012), which can be traced back to as early as Han Dynasty (206 BC–220 AD) (Yu & Jin, 2014; Carless & Lam, 2014). Historically, formal English language education and assessment in China started in the 1860s (Guoxing & Yan, 2014; Dello-Iacovo, 2009). The biggest nation-wide testing event in China is the National College Entrance Exams (NCEE), a summative method of selection for university entrance established in 1977 and which plays a vital role in a student's life because it is almost the only route to higher-paid jobs and higher social status (Carless & Lam, 2014). Moreover, success in summative assessment, especially in NCEE, can bring '*mianzi (face in Chinese which refers to how one wants to be treated and how one actually treats others along with their social-conception expectation* (Ting-Toomey, 2017)*)*' to the family. Due to the enormous influence of the NCEE, students' lives in schools are naturally examination-dominated, hence, assessment in education is underestimated in its scope and the term is sometimes used to refer to tests or examinations.

However, the idea of assessment came to the fore because of the criticisms of the intense focus on examinations, as well as the international influence of the widespread use of formative assessment. Because China has a top-down education system (Cherng, Hannum, & Lu, 2014), scholars and teachers are bound to implement what the government publishes. Examination-oriented teaching methodology gave way to other approaches under the guidelines of the *Basic Education Curriculum Reform Outline (Trial)* published by the Education Ministry in July 2001 (China Education and Reserach Network, 2005). Based on the reform outline, as well as the first New English Curriculum (NEC) in 2003, a new NEC (Ministry of Education, 2011) was adopted and published in January 2012. It includes an emphasis on formative assessment. One of the basic principles of the NEC is to recognise the vital role of formative assessment in promoting learners' development. The Chinese Ministry of Education define FA as making use of "self-assessment, peer-assessment, and assessment conducted by teachers and school administrators […] to observe, evaluate and monitor the learning process for the purpose of enhancing effective learning" (Chen, May, Klenowski, & Kettle, 2014). Moreover, it provides plenty of specific assessment techniques. In the appendices of the NEC booklet, scholars give several examples of FA in listening, speaking, reading and writing. From this point onward, FA has become increasingly popular in China.

At the meso-level, in Chinese academia more attention is now being given to FA than ever before. Between January 2001 and December 2012, there were only 36 articles published in the core journals in China in the field of foreign language research (Luo, Huang, & Xiaolei, 2014). Many of these studies merely mentioned FA's significance but did not study it. The practices relating to English assessment in China still remained under-represented (Guoxing & Yan, 2014) in research. Using the key term *formative assessment* (in English and Chinese), I searched the CNKI (China National Knowledge Infrastructure, known as the China Integrated Knowledge Resources Database, which covers all publications in journals in China). Only journals in SCI (Science Citation Index) and CSSCI (Chinese Social Sciences Citation Index) were searched because they apply rigorous quality and peer review criteria. This search found that 198 relevant articles were published in China over the 2001-2012 period (Table 2.4). On searching all the journals in the database, the total number of articles increased to more than 2000. In other words, only 10% of research studies on FA made their impact via published articles in high quality journals. However, the trend shows that there is a growing number of articles and increasing attention on the field of FA, especially in the years 2015 and 2016.

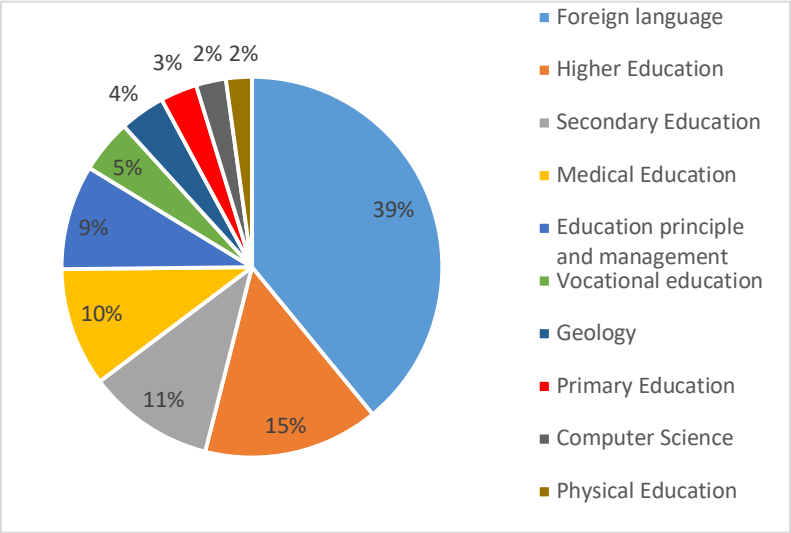
Table 2.4 Number of FA articles published in SCI and CSSCI in China

Year 20xx	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17
Number	3	6	10	11	8	15	17	28	21	22	25	32	26	28	40	41	26

If we extend the range of journals from 2001 to 2017, as shown in Figure 2.4, only 3% of the articles focus on primary education. However, in the period January to November 2017 alone, there are about 200 published journal articles in China on the theme of FA, evidencing strong growth in this area. It is clear that more FA research has been done, in English teaching and learning in particular. In April 2017, the First Conference on EFL Teaching and Assessment was launched in Beijing. For the first time, the Chinese academic community brought FA into public view as a key concept. Along with the conference, more

funding was released for FA research in secondary education and higher education. There will be greater opportunities and funding for the researchers who study FA.

Figure 2.4 Percentage of FA articles in different areas



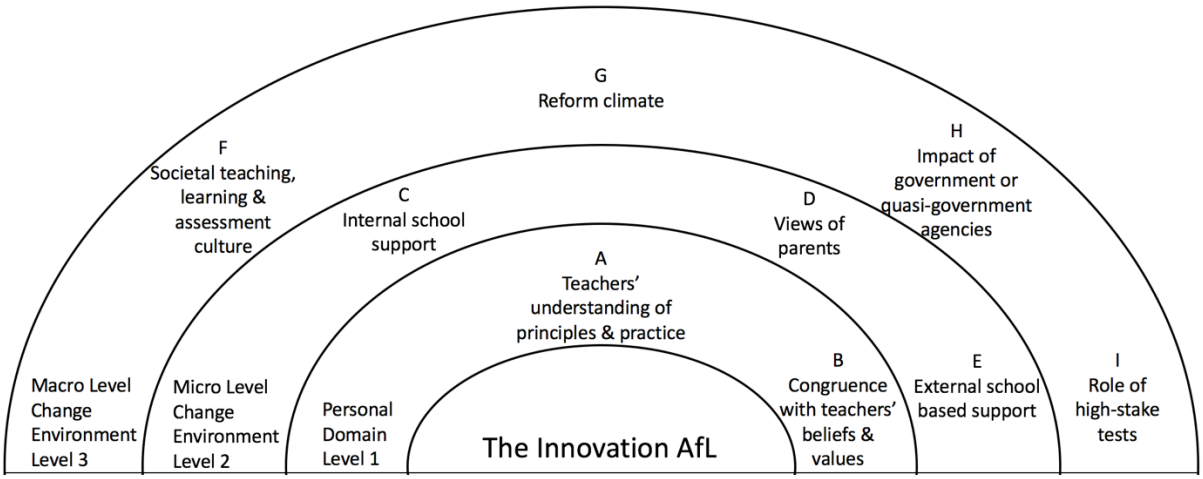
At the micro-level, many textbooks used in Chinese primary schools include a self-assessment form at the end of each unit. Based on the researcher's own observations of English teaching and learning in China (Zhang, 2013), self-assessment during the main activities of learning and teaching does happen but only infrequently. In the case study in Chen *et al.* (2014), it was found that English language college teachers in China did not actively engage students in peer-assessment or self-assessment. Self-assessment in China mainly relies on the self-assessment forms in the text books, rather than spontaneous activity. Figure 2.3 depicts a circle of learning units with FA. It is continuous since any assessment at the end of learning cannot be called formative because no further improvement could be made in the unit's time frame (Black & Wiliam, 2004). In this study, the FA strategies mainly referred to the pedagogical activities within the classroom, which did not include activities such as "Formative Use of Summative Testing" that brought up by Carless (2012). The core action in the teaching and learning process is taking steps to change the situation on the basis of the data collected in the classroom using the FA strategies.

The implementation of FA is not yet sufficient and effective in China (Peng, 2017; Liu & Xu, 2017). During NEC's implementation, scholars have found that teachers cannot get to the heart of FA while their priorities are focused on more homework or examinations. Traditional SA still plays the dominant role, since both parents and the education system judge students' performance on the basis of SA scores that are produced from 'objective' standardised tests (Gordon, et al., 2014; Carless & Lam, 2014). However, there is evidence that some FA activities, such as the Profile system, have been employed. This is a termly booklet that records a range of student data: scores from each exam in different subjects; distinction awards; teachers' comments; students' reflections and future plans for the next term. However, there is a significant conflict between the need for scores showing the products of learning and the need for students' growth evidenced by the learning process. Most Chinese teachers will only use assessment strategies that are examination-oriented, or for the purpose of examination preparation, rather than the learning process itself (Chen,

May, Klenowski, & Kettle, 2014). It is not only China, but also other Confucian heritage cultures, that find the re-contextualisation of FA difficult. For example, Vietnamese scholars have found that the examination traditions heavily influence their university students' learning. Accordingly, teachers there have to put extra work into localised FA and integrate it into their existing rules and norms (Pham & Renshaw, 2015).

Apart from the examination-dominated tradition, other factors play their parts in hindering FA implementation but more scholars in environments espousing Confucianism are finding 'indigenised' ways to implement FA in their systems. Hong Kong is a good example. Hong Kong is a very special place which combines traditions and culture of the East and the West and this brings a certain social adaptability. This is apparent in the implementation of FA in Hong Kong schools. Carless (2005) proposes a multilevel model of factors impacting FA in Hong Kong (Figure 2.5). There are three strands regarding the innovation of FA: teacher factors, micro-level school factors, and macro-level societal factors. Within the first level, there are two factors - Factor A: teachers' understanding of principles and practice, and Factor B: congruence with teachers' beliefs and values. At the second level, there is Factor C: internal support from management, Factor D: views of parents, and Factor E: external support from or collaboration with colleagues. At the third level, there is Factor F: societal teaching, learning and assessment culture, Factor G: wider reform climate, Factor H: impact of government or quasi-government agencies, and Factor I: wash-back from high-stake tests. Carless adds the role and contribution of students to the crucial factors, but the model does not include this due to it being from teachers' perspective (Carless, 2012).

Figure 2.5 Exploratory framework of factors impacting on the implementation of FA



Carless's framework provides a good approach to analysing the use of FA in China, in particular regarding the external environment. In this study, more attention was paid to the personal domain, which is Level 1 in Carless's framework. More follow-up work which focuses on Level 2 and 3 could be conducted based on this study. At this stage, the study can fill the gap in the framework concerning the missing factor about students' perspectives. Meanwhile, more studies about FA itself should be done to enrich the picture in the literature of implementation of FA in China. This is one of the reasons why this study was conducted. Furthermore, it is important to note that the time in one school lesson is limited in China, where the class timetable is highly structured. There are 40-50 minutes in a normal lesson. Of all the FA strategies, which ones are more effective in a limited timeframe and

also welcomed by teachers and students? This study therefore pursues a practical exploration of FA strategies in the classroom.

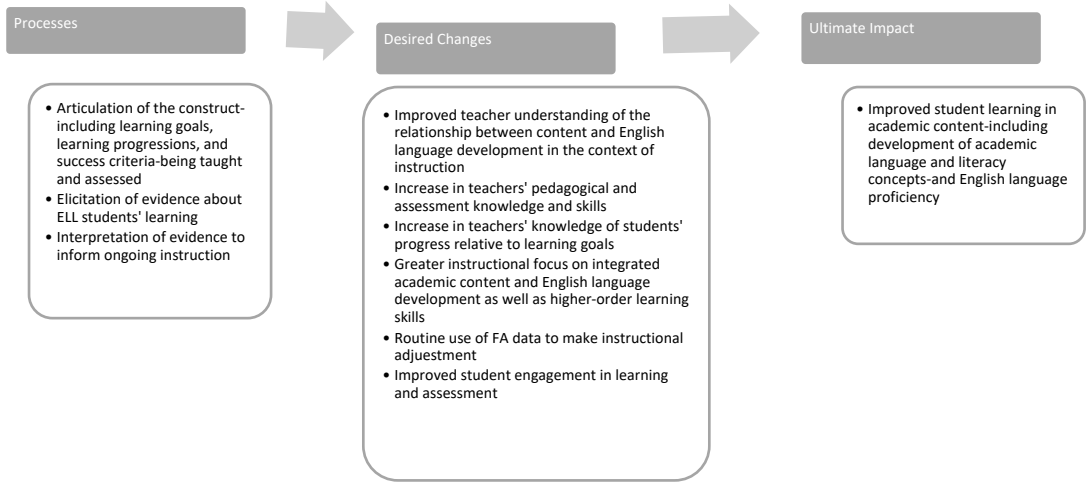
2.1.4 FA in language learning

FA plays different roles in different subjects. For example, FA in the mathematics classroom would be different for a history classroom, although certain common ground would also be share (Black & Wiliam, 2009).

As previously mentioned, during the last year (2017), 69.5% of the FA articles in China were actually located in the field of language teaching and learning. FA implementation in this field has had noticeable attention. It is probably because English is taught as a foreign language in China from year three (9-10-year-old) in most schools. As English is one of the three core subjects (the other two being Chinese and Mathematics), English language skills are one of the key areas that students have to master during their schooling. Even in Higher Education, the College English Test (CET) Band Four is a stepping stone which must be passed by every student in order to graduate. The CET has been a large-scale standardised English proficiency test in China since the mid-1980s (Chen, Kettle, Klenowski, & May, 2013). Therefore, FA in language learning has naturally become a hot area for attention.

Research has reported that language assessment strategies are naturally used in the classroom and they have formative potential for both teachers and students (Afitska, 2014a). FA strategies, specifically feedback, questioning, and peer-assessment and self-assessment, have a clear positive effect on students' linguistic development. Teachers and students show a positive attitude to these strategies too. Moreover, peer-assessment is a useful but risky approach (Butler & Zeng, 2014) because it is hard to achieve high levels of equality and mutuality in peers' interactions, particularly when a task places a heavy cognitive burden on young learners that may be beyond their abilities. Figure 2.6 depicts the logic model for the FA process (Alvarez, Ananda, Walqui, Sato, & Rabinowitz, 2014). There are three steps in this process: 1) the use of FA processes; 2) the changes after using FA desired changes; and 3) the FA's ultimate impact. The first step is similar to the use of FA in any education subject. Then the model posits that FA may lead to desired changes, such as students' engagement, or teachers' understanding, knowledge and skills. Lastly, the ultimate results or impacts, which are the phenomena I am exploring in this study, relate to the students' development in language proficiency.

Figure 2.6 Logic Model for a Proposed Approach to FA of English Learners



Adapted from Alvarez, Ananda, Walqui, Sato, & Rabinowitz (2014)

The framework above shows a good example of the use and extension of FA. From the perspective of long-term impact, FA aims to improve students' learning; the desired changes among teachers and students are the essential steps to achieve the ultimate goal. However, the implementation of FA is not elaborated here. Using terms like *articulation* or *elicitation* convey only the intention, not actual actions. However, the desired changes are valuable guidelines for analysing the teachers' perspectives on FA. The framework is more analytical than instructional for this study.

2.1.5 FA in teachers' perceptions

There is increasing evidence showing that the quality of the teacher is, if not the most important, then one of the most important determinants of educational outcomes of students (Lee & Wiliam, 2005), so it is important to find out how teachers see FA in their teaching. Despite the long-term promotion of FA around the world, teachers remain underprepared to use it in Hong Kong (Lam, 2015), not to mention the mainland China. There are many factors that can influence teachers' perceptions of FA. Many researchers have shown that teachers' beliefs, attitudes, perspectives and philosophies about teaching and learning are important in FA (Carless, 2005). Their purpose in teaching, whether to help students achieve more or to cover the content in curriculum, can affect the quality of the implementation of FA (Heitink, Van der Kleij, Veldkamp, Schildkamp, & Kippers, 2016). For example, in a recent study of teachers' perceptions of FA in China and Finland, the results suggest that Chinese teachers emphasise the positive impact of FA itself, while the Finnish teachers stress students' autonomy (Peng, 2017). It is also important for teachers to meet their personal goals reflecting their values (Harrison, 2005). They are enthusiastic about their personal professional development when applying FA in their teaching, and when reflecting on it (Black & Wiliam, 2004). Black and Wiliam encourage teachers not to work harder, but to work smarter (ibid).

When it comes to the use of FA, some of the strategies are used more often than others, which implies the uneven understanding or implementation of FA (Wylie & Lyon, 2015; Lysaght & O'Leary, 2013). Of course, this could be because of the nature of different FA strategies, given that some strategies, such as questioning, are naturally embedded in the

whole teaching process. But in some cases, teachers will use FA in their preferred ways which do not match the actual definition. For example, teachers may purport to use 'FA', but in fact be carrying out a process assessment instead, which emphasises timing and frequency using multiple mini-tests during the learning process, for the purposes of reporting (Rea-Dickins & Gardner, 2000; Chen, 2011; Chen, Kettle, Klenowski, & May, 2013). The flexibility enables teachers to apply the form of FA in their own teaching, *i.e.* collecting information consistently, yet they do not get to the heart of FA, *i.e.* using the information collected to adjust teaching. This problem echoes the issues about localisation of FA during implementation that will be discussed in 2.1.9.

Moreover, according to Black *et al.* (2003), teachers refuse to separate FA and SA, and they found the formative use of summative assessment to be useful. The traditional ways of assessment still play an enduring role even though FA is introduced. Teachers feel insecure and that they are 'losing control' in their teaching (Lee & Wiliam, 2005). Heitink *et al*. (2016) recommend that teachers study the constructivist view of learning and pedagogical strategies. It is not only their idea of FA, but their whole set of teaching beliefs which could be reviewed and renewed. Teachers' understanding of FA is deepened when constructive discussions happen between them and their peers. There is evidence that the quality of teaching improves over the first five years, then levels off, and then declines according to some studies (Hanushek, Kain, O'Brien, & Rivkin, 2005). In this study, there are two teachers involved: one has rich experience, including using plenty of FA; while another is a novice teacher with two years of teaching experience and little experience of using FA.

It is not only their understanding of assessment, but also their working environment that affects teachers' perceptions. Although some teachers report feeling empowered by using alternative assessment despite the resistance from faculty and the extra workload (Carless, 2005; Lam, 2016), many other teachers find FA to be time-consuming partly because FA practices are mostly teacher-originated. This leads to high workloads for teachers (Jonsson, Lundahl, & Holmgren, 2015; Black, 2015). In China especially, because the pupil-to-teacher ratio is astonishingly high, even a slight change in assessment habits will cause significant extra work. However, as aforementioned, if the teacher has a supportive working environment with professional development practices that meet teachers' needs, teachers are happy to take up FA practice (Lee & Wiliam, 2005).

To sum up, at the level of individuals, teachers' beliefs and attitudes are significant for FA implementation; at the level of working environment, supportive and constructive communication is needed among teachers. Teachers are pulled back and forth by the imperatives for personal development that teachers value, on the one hand, and extra workload due to the changing assessment habits, on the other. They do not want to lose control in the classroom, but meanwhile are looking for new strategies that can be effective in teaching.

2.1.6 FA in primary education

In China, the Ministry of Education has released *College English Curriculum Requirements,* which recommend the juxtaposition of FA and SA (Guoxing & Yan, 2014), similarly to the NEC which is applied in primary and secondary education. Therefore, the use of FA is

mandated throughout the education system. Nevertheless, compared to higher education, FA research in primary education is under-represented (Carless & Lam, 2014). The research studies that have been conducted among young learners are few in number, especially in respect of 8 to 10-year-olds (Afitska, 2014b).

Many well-known assessment projects and protocols which have had a significant role in promoting FA around the world have been implemented in schools, but mainly in secondary schools. For example, the *King's-Medway-Oxfordshire Formative Assessment Project (KMOFAP)*, led by Black and Wiliam *et al*. (2002; 2003) was applied in six schools, involving students age 11-18 years and 48 teachers. There are more examples in recent years, such as the four-year *English Learner Formative Assessment (ELFA) Project* from 2010 in the USA (Institute of Education Sciences, 2010), and a two-year *Embedding Formative Assessment (EFA) Project* involving 70 secondary schools from 2015 in the UK. Although each has a different framework, all the projects confirm the significance of FA in teaching and learning but they report mixed feedback from schools, teachers, and students.

There is however some research that has focused on the primary school context, for instance the *Teacher Assessment at Key Stage (TASK*) *Project* (Torrance & Pryor, 1998) and *Primary Response Project* involving 11 teacher-researchers in primary schools in the UK and led by Torrance and Pryor (2001); and the two-year *Primary English Assessment for Learning (PEAL) Project* involving 16 teachers from 10 schools in Hong Kong, led by Carless (2007), *etc*. Harris (2007) reports the use of FA in eight schools across the UK. Strategies such as target-setting by students, sharing the learning objectives, questioning, feedback by the teacher, and self/peer-assessment have been implemented, and bring achievement in meaningful learning. In light of the published research, it is possible that younger learners are more sensitive to FA than older learners, and the impact of FA is potentially more significant.

All of the projects referenced above were long-term with at least a one-year procedure. Unfortunately, no short-term (less than six months) research studies have been found in this area, possibly because of funding institutes' preference for longer-term research as well as the nature of assessment practices which may take a longer time to embed in the classroom and for the effects to be manifested by teachers and students. However, due to the practical constraints on the researcher explained further in 3.8.2, this study investigates the impact of FA in a short-term context.

2.1.7 FA with motivation and achievement

It is said that FA can have a powerful impact on student motivation and achievement (Cauley & McMillan, 2010). This is one of the reasons why this study was conducted. When teachers share learning aims with students, the students can feel some 'ownership' of learning (Curry, Mwavita, Holter, & Harris, 2016). According to the Goal Orientation Theory in motivation, there are two types of achievement goals: mastery goals and performance goals (Ames, 1992). Mastery goals refer to the intention to master and understand content, and the willingness to engage in the learning process; performance goals indicate that one can demonstrate their ability by outperforming others, or by achieving success with little effort. It is said that students tend to internalise the mastery goals if FA is used properly, which accordingly leads to intrinsic motivation (Cauley & McMillan, 2010). Students are

given a voice in setting challenging goals, which actively involves the student in the learning and enhances their motivation (Curry, Mwavita, Holter, & Harris, 2016; Harris, 2007). Students' attitudes towards learning can be shaped by feedback from a very early age (Wiliam, 2011). Students are willing to learn more if they can 'see' that their aims are achieved. Therefore, the type of feedback will influence the students' motivation, particularly in lower achievers since their previous experience may have led them not to try very hard (Black & Wiliam, 1998b). Self-assessment or peer-assessment helps develop students' autonomy when they are aware of the assessment criteria or mark scheme (Harris, 2007).

The positive effect of FA on achievement is a matter of debate. A recent study about the impact of formative assessment on Year 4 student achievement shows that students with controlled pre-test scores significantly outperformed other students in a post-test after a school year (Andersson & Palm, 2017). It is not only primary school students, but also college students who show better achievement levels if they participate in FA (Carrillo-de-la-Peña, et al., 2009). Based on an influential review of 250 articles between 1987 and 1997 (Black & Wiliam, 1998a) as well as their *KMOFAP* (Black & Wiliam, 2004) and a study of the impact of FA on the achievement of secondary school students (Wiliam, Lee, Harrison, & Black, 2004), it was concluded that FA did have a positive impact on students' achievement in the quantitative results of high-stake examinations in the UK.

However, this much-cited review's finding of significant improvement in achievement has been questioned by Dunn and Mulvenon (2009) and other researchers. Dunn and Mulvenon state that most of the articles in Black and Wiliam's review are lacking generalisability. Even though the 1998 article has been cited more than 8000 times, there is still a need for more empirical evidence into the effects of FA. Sometimes, FA strategies could be more effective if they are embedded in each other (*e.g.* feedback and goal-setting) (Kluger & DeNisi, 1996). Moreover, Bennett (2011) claims that Black and Wiliam's review article is based on limited sound scientific evidence. But there are several studies that have found that FA has no to little impact on learning (Heitink, Van der Kleij, Veldkamp, Schildkamp, & Kippers, 2016). For instance, it is reported that assessment as learning in the portfolio-based context may not necessarily improve students' writing achievement, despite the students' positive attitude to it (Lam, 2016).

This research aims to contribute a solid empirical study of FA and offer more evidence of the impact of FA on students' motivation and achievement. It also aims to enrich the corpus of FA studies in Anglophone literature with the under-represented culture context: China.

2.1.8 FA with technology

Even though much FA research has been conducted in traditional classrooms, FA with technology has also been thoroughly studied (Dunn & Mulvenon, 2009). The EFA project recommends more ICT-based technology and software/apps to enrich feedback and allow individuality to flourish (SSAT, 2017). Gordon, *et al*. (2014) also introduce an example of classroom practices which involve technology. The students use an online system on tablets to discuss and debate, the teacher makes notes instantly and records the results in a Venn diagram, and uses questioning of students' responses to lead them into deeper thinking.

Additionally, the teacher can give private feedback via the system to a particular student while physically approaching him or her when students articulate the conclusion in the large group. Feedback sometimes has a negative impact on students, lower-achievers in particular, for making the students believe that they lack ability to study (Black & Wiliam, 2004; Black & Wiliam, 1998b). In this example, the teacher has rich information with which to understand the students' progress and to improve the next steps of teaching and learning. Teachers can even review the recordings of the class online after the lesson and learn from past mistakes. Therefore, teachers can give synchronous and asynchronous feedback, such as 'real-time' adjustments in the classroom and 'follow-up' comments after the class (Black & Wiliam, 2009). This is hard to achieve in a traditional class where teachers cannot observe every student. The pressure to value every contribution is strong, since a student might feel rejected if his or her contribution is dismissed. With technology, real-time adaptations in teaching and learning are possible. Also, data-driven decision-making can reduce workload and even reduce the pressure on teachers (Luckin, Clark, Avramides, Hunter, & Oliver, 2016). Although FA with technology is not the key focus of this study, it is a promising research direction in the future and worth signposting here.

2.1.9 Issues for implementation of FA in China

The development and implementation of curriculum and assessment needs time; professional development and practical tools are needed for practitioners (Gordon, et al., 2014). Teachers accept new methods only infrequently often due to their impracticality. They prefer methods tried and tested by themselves (Torrance & Pryor, 2001; Brown & Mcintyre, 1993). That is why Torrance and Pryor recruited teacher-researchers in the *Primary Response Project*. However, in the implementation of FA, the people who will use FA strategies are front-line teachers. Some of them may have been teaching for years and will have their own established teaching habits.

Moreover, front-line English teachers' needs do not always match the academic findings in studies. Firstly, classes in Jiangsu Province in China normally have 50-60 students in each. An English teacher usually undertakes teaching duties in two to four classes. With the heavy workload in primary school, English teachers in China do not have much time to study the theories behind educational concepts; they may find FA, as a new concept, difficult to digest. However, because of the top-down nature of policy directives in China, the format of FA will inevitably be learned but the core rationale could be neglected. The contextual and local cultural differences are underestimated before enacting new systems (Chen, May, Klenowski, & Kettle, 2014). There is an analogy with training for international competitions, where some students are coached to achieve very high scores but, paradoxically, they develop only limited or partial understanding of what they are assessed on (Carless & Lam, 2014). Secondly, the conception of FA is vague, even for the researchers who are studying it (Taras, 2010). The FA is impracticable when it comes to external implementation beyond the academic group. This is the reason why I explain the FA principles in detail in 2.1.1. When talking to school-based English teachers at a conference, they complained that the 'theoretical talks' were useless to them, but that reports with actual exam results were helpful. Many research studies lack the details that teachers can use during implementation (Black & Wiliam, 2004). Teachers prefer direct instructions and transparent guides that tell them which strategies are effective in teaching and learning and they are happy to use them

accordingly. Therefore, what teachers need are straightforward concepts and handy tools which can be applied in their classrooms immediately, but there is a danger of this leading to a certain degree of stereotyping and over-simplification. The right balance between these competing needs is difficult to achieve.

Students' responses to FA in Confucian-influenced culture differ from those in western countries. The collectivism in China, as well as the hierarchical system that requires respect for teacher authority, results in students who are less rebellious than in many other cultures (Carless & Lam, 2014). Students are used to the mono-directional relationship with teachers, which can cause difficulties in peer-assessment and self-assessment. In addition, the ingrained habits of SA in Chinese schools and the deep-rooted belief in the importance of instant success mean that prospects for the uptake of FA are uncertain. The aims of this study is to find out the most effective way of using FA strategies that applied in primary school Chinese context from the perspectives from both teacher and students. Although FA has been implemented for a decade, not much work has been done to compare the effectiveness of different strategies.

2.1.10 Summary

As Dylan Wiliam has said, the idea that assessment can support learning is not a new idea. However, it is only 47 years since Benjamin Bloom first brought the concept of FA into education; and only 20 years since Paul Black and Dylan Wiliam drew wider public attention to it; not to mention that it is only 15 years since the first book about FA in English teaching and learning was published in Chinese and introduced FA systematically to China for the first time. FA is still an area that needs to be explored and validated in the context of China.

In the literature review on FA, the fundamental concepts such as assessment and measurement/evaluation, FA and SA were defined and differentiated. Formative assessment, as an ongoing assessment tool in teaching and learning, aims to improve students' learning by collecting information about their learning using five strategies and thereupon adjusting teaching on the basis of the new information. Three FA frameworks were introduced and elaborated. Combining the key strategies that all models have in common, an FA framework was constructed for later application in this study, comprising the following five strategies: sharing learning aims, questioning, feedback, peer-assessment and self-assessment. A key focus was placed on the implementation and drawbacks of FA in China, as the culture difference influences the use and effects of FA in teaching. Although the Chinese government has emphasised the need for FA in English teaching, and more and more FA studies have been conducted, the practice is still heavily affected by the SA-dominated tradition and there are many more hindrances in applying FA in China. The present work also considered and discussed the specific contexts of this study, for instance, FA in language learning, in primary education, teachers' views of FA, and the connections to motivation and achievement. The debate around the impact of FA on motivation and achievement is ongoing so this is a good opportunity for the researcher to provide new empirical evidence to contribute to the FA literature. Lastly, it was important to note that as technology develops, FA will not be limited to traditional class contexts and methods. A short paragraph in this chapter has addressed the potential of technology-enhanced FA in the future.

In the next section, the literature about motivation, especially the concepts, theories, and frameworks for motivation in language learning, will be reviewed.

2.2 Motivation

2.2.1 Defining Motivation

It is widely accepted that motivation is one of the key factors in language learning which drives the learning process (Oxford & Shearin, 1994; Dörnyei, 1998a). Motivation is a complex concept that can be interpreted in many ways. In the Oxford English Dictionary, the term 'motivation' means "a reason or reasons for acting or behaving in a particular way" or "desire or willingness to do something". Nevertheless, motivation can refer to something more particular when it comes to the field of language learning. Both motivation and language learning are complicated phenomena with many factors and variables, not to mention when they are combined. However, there are some commonly agreed components in the concept of motivation: the choice of a particular action; the persistence with it; the effort expended on it (Dörnyei & Ushioda, 2013).

Gardner, who brought the study of motivation into the field of second language (L2) learning for the first time, defines motivation in L2 learning as "the extent to which an individual works or strives to learn the language because of a desire to do so and the satisfaction experienced in this activity" (Gardner, 1985: p.10). He also points out that motivation concerns "those factors that energise behaviour and give it directions" (p.281). Pintrich and Schunk (1996) define motivation as "the process whereby goal-directed activity is instigated and sustained". They see motivation as a process, rather than a "fairly static mental or emotional state, or a goal" (Dörnyei, 1998a). Therefore Dörnyei includes the idea of process-oriented perception in a synthesis of the static and dynamic conception of motivation, summarised as: "a process whereby a certain amount of instigation force arises, initiates action, and persists as long as no other force comes into play to weaken it and thereby terminate action, or until the planned outcome has been reached" (Dörnyei, 1998b). Williams and Burden elaborate the definition as: "motivation may be construed as a state of cognitive and emotional arousal, which leads to a conscious decision to act, and which gives rise to a period of sustained intellectual and/or physical effort in order to attain a previously set goal(s)" (Williams & Burden, 1997).

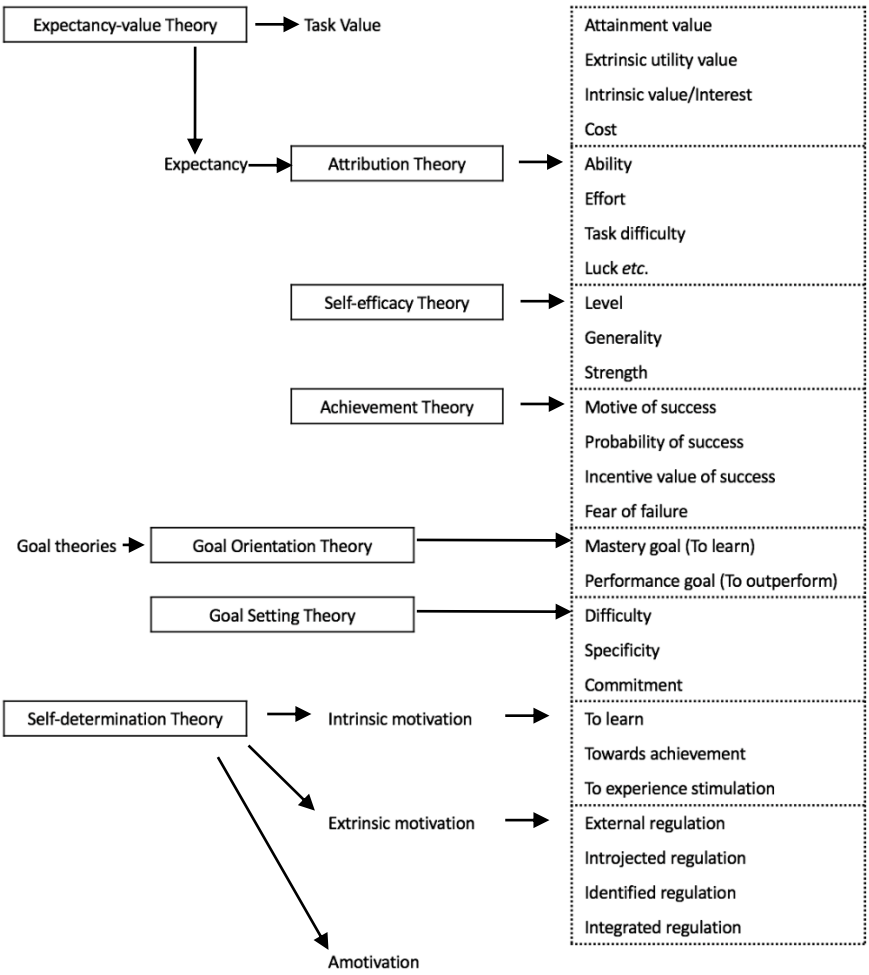
Both Dörnyei and Gardner agree that motivation is "one of the most elusive concepts in applied linguistics and indeed in educational psychology in general" (Dörnyei, 1999), and it has different meaning to many different individuals (Gardner, 2001). In this study, motivation is operationally measured as the learners' feedback via the motivation questionnaire designed by the researcher on the basis of various motivation frameworks. The main focus on motivation here is on a general measurement which can reflect the degree of motivation in different dimensions, rather than a focus on one particular motivation theory. It is true that each motivation theory has its own strengths. However, to customise the motivation measure for this study, a single motivation theory is not enough. As an example, in sharing learning aims, the type of the aims (whether mastery or performance) and the features of aims (whether difficult or easy, specific or vague, *etc.*) are

directly related to Goal Orientation Theory and Goal Setting Theory. They are irreplaceable by other motivation theories as none of the other theories mentions aims - an essential element in this study. After combining different aspects of motivation, its definition for this study is: “the beliefs, values and goals that determine which tasks students pursue and their persistence in achieving them, and the standards they set to determine when a task has been accomplished” (Wentzel & Wigfield, 2009).

2.2.2 Motivation theories in psychology

Motivation plays an essential role in the educational psychology field. Because of the vibrant development of motivation theories, many people will be confused about the complex relationships between the theories. At first, motivation theories emphasised the drives and needs of individuals as well as the rewards and punishments they received. Later, social cognitive theories dominated the field, hence the notions of individuals’ beliefs, attributions, and sense of control over efforts started to move to the front of the stage (Wentzel & Wigfield, 2009). To make it simple, I map out the current mainstream motivation theories with their key variables in Figure 2.7. As the theories are compatible with each other due to the similar variables, it is important to reorganise the theories rather than merely list them. Although all the variables in each theory are arranged and depicted in one figure, they do not function on the same level. Therefore, another paragraph will present a synthesis of the seven theories in a way that makes more sense having introduced each one.

Figure 2.7 Construct of Motivation Theories



Although the figure is simplified and clear, the motivation theories are more complex than a few variables. The only purpose of breaking down and listing the elements is to help make motivation measurable via its components. There are more details requiring discussion here, so I will elaborate each theory in the following section, and in 2.2.3 Motivation frameworks in language , I will fit all the theories into the motivation framework for the current study, with the help of motivation models of language learning.

Expectancy-value Theory

Expectancy-value Theory in motivation is "the most long-standing and influential framework in cognitive motivational psychology" (Dörnyei, 2001; Dörnyei & Ushioda, 2013). Behaviour is a function of the expectancies one has and the value of the goal toward which one is working. Atkinson first introduced this theory (Atkinson, 1957; 1964) and at the time he defined *expectancy* in an operational way as the proportion of individuals who have succeeded at the task in the past. Later, Eccles and her colleagues (Eccles-Parsons, et al., 1983; Eccles & Wigfield, 2002) defined *expectancies for success* as "individuals' beliefs about how well they will do on upcoming tasks, and ability beliefs about how good one is", while *values* are the reasons why individuals do the tasks or activities: is the task important, useful or interesting? Pekrun identified three expectancy beliefs in his Control-value Model (Pekrun, 1993; 2009): situation-outcome expectancies - a situation that will produce an outcome; action-outcome expectancies - one's beliefs about the consequences of their

actions; and action-control expectancies - one's belief about whether they can do a certain action. In all the expectancies, action-control is similar to self-efficacy that Bandura proposed in his work (1997).

Several theories focus on the expectancy component of the expectancy-value dichotomy, such as Attribution Theory, Self-efficacy Theory, Achievement Theory, *etc*. (Atkinson, 1957; Bandura, 1986; Eccles & Wigfield, 1995). Eccles *et al.* (Eccles-Parsons & Goff, 1978; Eccles-Parsons, et al., 1983; Eccles, 1987) also build their Modern Expectancy-Value Theory with a richer picture of the 'value' component. Because value regards the individuals' desire to do the task, the value component is referred to as 'task value' (Eccles, 2005).

Eccles-Parsons and colleagues (Eccles-Parsons, et al., 1983) developed four major components in value, which are: attainment value/importance, extrinsic utility value/usefulness of the task, intrinsic value, and cost. Attainment is the individual's thoughts about the importance of doing well in a given task. It shows one's confidence in identifying issues, because a task is important to them when they view themselves as the central to their own sense of themselves (Wigfield, Tonks, & Klauda, 2009). Utility value is sometimes seen as the extrinsic motivation because the tasks are done out of a sense of the utility purpose, a path to the end rather than an end itself (Ryan & Deci, 2009). It also can be related to intrinsic motivation, where people want to be 'a better me' via certain tasks. Intrinsic value here refers to the enjoyment of doing a task, which is very similar to intrinsic motivation in self-determination theory. These three values ask three main questions respectively in task value: is this important to me?/ is this useful to me?/ is this interesting to me? They are all 'positive energy' in task values, while 'cost' is the only critical part. Cost is about people's loss in doing a task. Individuals may ask themselves: what do I need to give up/offer if I participate? Expectancy of success and values are inseparable because the level of expectancy affects the value, and *vice versa*.

Expectancy-related beliefs and values decline when students grow older, and the beliefs and values are varied in different subject areas. The expectancy beliefs and values decline intensively in language, maths and sports when the individuals are in primary school (Wigfield & Cambria, 2010). One of the explanations is the over-emphasis on evaluation which leads to students' de-valuation of subjects they do not do well (Wigfield, Eccles, & Rodriguez, 1998). Therefore, the changes in optimistic and realistic feelings, and even pessimistic feelings, are particularly important when motivation is studied in young children.

Attribution Theory

Previous activities must have some influence on current or future tasks (Dörnyei, 2003). Attribution Theory is the one which concerns individuals' interpretation of their past achievement outcomes to determine subsequent behaviour (Eccles & Wigfield, 2002; Kelley & Michela, 1980). This theory is originated from Fritz Heider (1958). He believed that people are amateur scientists who keep putting information together until they find a reasonable explanation or cause. It was then developed by Bernard Weiner (1986) and colleagues with their elaborated theoretical framework. Once people find the means to success, they may attempt to repeat the same procedure to achieve success again; likewise, if they know the

reason why they fail, they could stop doing the same thing to avoid failure (Weiner, 1972; Kelley & Michela, 1980).

In Attribution Theory, causal attributions, as a mediating link, connect past experience to future choices (Keblawi, 2006). Generally speaking, former failures leads to abandonment, while successes encourage people to try again. For example, if a student failed in a question and received negative feedback from the teacher, when it comes to a similar question, the student may think carefully before answering. In contrast, if the student answered a question successfully, s/he will try to answer a similar question better next time because success seems easy to achieve and surpass. In this ongoing context, because of the pride of success and the shame of failure, people have a strong tendency for successes, which matches the core idea in Expectancy-value Theory (Weiner, 1992).

People can use Attribution Theory to answer 'why' questions. For instance, to answer the questions 'why I failed the exam', the answers could be 'I am not smart' or 'I did not work hard'. Such causal thinking about the explanations or reasons for phenomena happen spontaneously in people's everyday life. Frieze and his colleagues (Frieze, 1976; Frieze & Snyder, 1980; Elig & Frieze, 1980) identify several dominant attributions of success and failure, such as effort, ability, task, luck, interest, and other persons. Weiner (1992) also proposed the four most important achievement attributions - ability, effort, task difficulty, luck - and classified them into three causal dimensions, as shown in Table 2.5. It illustrates the relationship between the causal dimensions and causes of success. Three causal dimensions are identified by Weiner (Weiner, 1986; 1979), which are: locus (esteem-related affect) – whether a cause is internal or external to the individual; stability (expectancy) – whether a cause is constant or varying over time; and controllability (social emotions) – whether a cause can be influenced easily. Theoretically, all the causes can be labelled with the three dimensions (Graham & Williams, 2009).

Table 2.5 Causes Ascriptions Related to Causal Dimensions

	Causes			
Causal dimensions	Ability	Effort	Task difficulty	Luck
Locus of control	Internal	Internal	External	External
Stability	Stable	Unstable	Stable	Unstable
Controllability	Uncontrollable	Controllable	Uncontrollable	Uncontrollable

Self-efficacy Theory

According to Bandura (1988), the most important expectancy that we learn lies in Self-efficacy Theory, a motivation model which focuses on the performance capabilities rather than personal qualities. Pintrich *et al*. (Pintrich, Smitch, Garcia, & McKeachie, 1991) define self-efficacy as "a self-appraisal of one's ability to perform a task, and it contains one's belief and confidence in oneself to achieve that specific task". Bandura (1997) also defines self-efficacy as an individual's belief/judgement/confidence that s/he has the capability to reach a certain level of performance or attain a designated goal. He also postulates the sources of

self-efficacy information, which are actual performance, vicarious experience, forms of social persuasion and physiological indexes (ibid). In all the sources, the actual performance is the most reliable way to assess one's self-efficacy (Schunk & Pajares, 2009). However, these sources do not have direct impact on people's judgements. It is the type of information and the rules people use to integrate the information that form the basis of interpretations. The judgements of self-efficacy can be influenced during the process of selection, integration, interpretation and recollection of information (Schunk & Pajares, 2009).

This "multidimensional construct" (Bandura, 1997) varies in three determinants: strength, generality and level/difficulty. These components differ in form and function. Strength of efficacy shows the level of certainty of accomplishing a given task; generality is whether the confidence can be shared among different activities (transferability); and difficulty of different tasks can affect the level of self-efficacy as well. For example, an individual's quantifiable confidence in easy mathematics might be 100 per cent, but s/he may become 80 per cent confident when s/he meets difficult mathematics questions or 50 per cent confident in language learning. Consequently, even if individuals have the ability or the skills to perform successfully, they still need strong beliefs that encourage them to participate more positively and contribute more in different tasks. Here, it is important to differentiate self-efficacy from self-esteem, as they are two different concepts. Self-esteem is about 'how do I feel about myself' as a writer/researcher *etc.*; while self-efficacy is 'am I capable of writing a PhD thesis'. However, in this study, the participants are school students, whose sense of themselves and their abilities may not be well-calibrated. It is common for them to overestimate themselves due to the unapprehended demands; or to underestimate themselves due to inappropriate instructional and social factors (Schunk & Pajares, 2004). Pajares (2003) found that in language learning, girls' self-efficacy is similar to boys', even though girls' achievement is higher.

An individual's efficacy expectation is the major determinant of goal-setting, activity choice, willingness to expend effort, and persistence, which links Self-efficacy Theory with other motivation theories, for example: Expectancy-value Theory, Goal theories, Self-regulation Theory (Hsieh, Sullivan, & Guerra, 2007; Zimmerman, 2000). Zimmerman (2000) describes this as "an essential component for successful learning". Nevertheless, Bandura (1997) states that self-efficacy is not the only influence on behaviour, because a person who is highly self-efficacious does not necessarily have strong ability.

Achievement Theory

Most motivation theories are rooted in achievement context, for instance: Expectancy-value Theory (*e.g.* Wigfield & Eccles, 2000), Self-efficacy Theory (*e.g.* Zimmerman, 2000), Attribution Theory (*e.g.* Graham, 1991), Goal theories (*e.g.* Elliot & Church, 1997), Self-determination Theory (*e.g.* Pae, 2008), Self-regulation Theory (*e.g.* Zimmerman, 1990), *etc*. Although the achievement context is almost ubiquitous in motivation models, it cannot be taken for granted or neglected.

The concept of achievement motivation and its systematic approach were initially introduced in the mid-20th century (Atkinson, 1957; Atkinson, 1964; Atkinson, Raynor, &

Birch, 1974; McClelland, 1967). This theory is based on how important the individuals' experiences are and how they work to approach a satisfactory performance. It focuses on four factors, which are: the need for achievement (the motive for success); the probability of success (expectancies of success); the incentive value of success; and the fear of failure (Oxford & Shearin, 1994; Graham & Weiner, 1996; Dörnyei & Ushioda, 2013). When approaching an achievement-related goal, L2 learners are motivated by the probability of success and the incentive value of success. Success brings satisfaction, whereas failure leads to shame. This approach-avoidance tendency is mainly considered in achievement goal studies. According to Elliot & Thrash (2001), it is a basic and fundamental step to tell the difference between these two conditions – approach and avoidance – in order to understand achievement-motivated behaviour.

Goal Orientation Theory

Goal Orientation Theory is sometimes referred to as Achievement Goal Theory. It was pioneered by Dweck (1986) and Nicholls (1984). There are many types of goals. The first perspective on goals concerns the individual's aim in a particular task or problem, hence, the task goal. The second perspective on goals is generated from an individual's general motive, such as why an individual is taking an action to do something. Unlike the task goals, the general goals do not necessarily have specific standards or criteria of success. They can also be applied to different contexts, rather than just achievement contexts. The third perspective on goals is achievement goals, which sit between task and general goals. It refers to why an individual is pursuing a task. Normally the context will be academic achievement tasks (Pintrich, 2000).

Table 2.6 illustrates the main taxonomy in achievement goals. Nicholls and his colleagues (Nicholls, Cobb, Wood, Yackel, & Patashnick, 1990) define two main branches in goals: ego-involved goals and task-involved goals. These are similar to what Dweck (1991) proposes on performance goals and learning goals. Ames (1992) also introduces performance goals and mastery goals. These goals are assumed to reflect the organised system or theory behind the activities. Therefore, the goal orientation is used here to represent the beliefs about "purposes, competence, success, ability, effort, errors, and standards" (Pintrich, 2000). The different goals form the basic construct in Goal Orientation Theory.

Table 2.6 Main Categories in Achievement Goals

Nicholls *et al*. (1990)	Ego-involvement goals Be superior to others	Task-involvement goals Gain understanding
Dweck (1991)	Performance goals Gain positive/Avoid negative judgement of competence	Learning goals Increase competence
Ames (1992)	Performance goals To demonstrate one's ability which evidenced by outdoing others, by surpassing normative-based standards, or by achieving success with little effort.	Mastery goals Master and understand content and demonstrate a willingness to engage in the process of learning

Conceptually, task-involvement goals, learning goals and mastery goals are quite similar. Learners with these goals all have a belief that the learning is concerned with their own

progress rather than being better than others. They want to 'become smart' themselves, rather than to 'look smart' in others' eyes (Öztürk, 2012). Mastery goals are helping individuals to develop competence, while performance goals attempt to demonstrate the competence. Ego-involvement or performance goals pay more attention to the comparison between 'me' and others. The two orientations fit the extrinsic utility value and intrinsic value in task values. Butler (1989) and Deci & Ryan (1991) encourage the integrating of Goal Achievement Theory with Self-determination Theory, due to the corresponding relationships of intrinsic/extrinsic motivation in Self-determination Theory and mastery/performance goals in Goal Orientation Theory.

There are not only the different goal orientations, but also different potential states. Elliot (1999) suggests dividing performance goals into two types: performance-approach goals, where learners emphasise "gaining positive judgement and beating others"; and performance-avoidance goals, where students tend to "hide the lack of ability". Pintrich proposes these two states within both goals (Table 2.7). Therefore, when pursuing a task, individuals may have different attitudes toward their achievement tasks due to their various results of attributions, efficacy, persistence, and choice *etc*. A good example is perfectionists: if they want to master knowledge in an academic context, avoidance of mistakes might be more important to them than truly understanding the tasks, since they can barely tolerate mistakes.

Table 2.7 Two Goal Orientations and Their Approach and Avoidance States

	Approach State	Avoidance State
Mastery goal	Focus on mastering task, learning, understanding	Focus on avoiding misunderstanding, avoiding not learning or not mastering task
	Use of standards of self-improvement, progress, deep understanding of task	Use of standards of not being wrong, not doing it incorrectly relative to task
Performance goal	Focus on being superior, besting others, being the smartest, best at task in comparison to others	Focus on avoiding inferiority, not looking stupid or dumb in comparison to others
	Use of normative standards such as getting best or highest grades, being top or best performer in class	Use of normative standards of not getting the worst grades, being lowest performer in class

Adapted from Pintrich (2000)

However, Goal Orientation Theory only answers how or why individuals are trying to achieve goals, but not what they are attempting to achieve (Kaplan & Maehr, 2000). In Goal Setting Theory, the content of the goals will be specified.

Goal Setting Theory

After understanding the difference between goal types, Goal Setting Theory reveals how to set goals effectively. This theory was first created to improve work performance (Locke & Latham, 1990). Later, it was applied in educational settings. Individuals who possess mastery orientation may choose high-level goals while those who possess performance orientation would prefer easy tasks as they do not want to take the risk of failure or be judged (Locke & Latham, 2006); this matches the approach-avoidance tendency in achievement-motivated behaviour, as mentioned earlier. It is also very likely that the self-efficacy of people and their performance can be promoted when they choose challenging goals (Bandura, 1997;

Schunk, 1990). According to Öztürk (2012), it is important to break tasks into sub-tasks with due dates, and to control the learning process gradually. This process is comparable with the internalisation of extrinsic motivation to intrinsic motivation in Self-determination Theory.

There are three basic features of goals in Goal Setting Theory: difficulty, specificity and commitment. It has been proved that "specific, high (hard) goals lead to a higher level of task performance than do easy goals or vague, abstract goals" (Locke & Latham, 1990; 2002; 2006; Locke, 1996). To make sure the goal works, learners need to commit to the goal whoever sets it; be able to reach the goal; receive some goal-relating feedback. Without goals, feedback is meaningless; without feedback, the person does not have a clear direction towards goals. These goal features are usually recommended to people who are procrastinating. Rather than setting big and vague task goals, it is more realistic to set doable, reachable goals. The difficulty of goals is lesser when the goals are smaller and more specific. The sense of achievement as well as the prompt goal-related feedback can promote the learning process. Goal-setting and teachers' feedback are the main elements in formative assessment, and this is the reason why Goal Setting Theory is considered in this study.

Self-determination Theory

Last but not least, it is vital to elaborate Self-determination Theory as it has the most well-known motivation categories: intrinsic and extrinsic motivation. Intrinsic motivation (IM) focuses on behaviour *per se*, and is founded upon inherent human needs for competence; while extrinsic motivation (EM) deals with incidental rewards or punishments and seeks an instrumental end. Nevertheless, there is another category of motivation: amotivation, which refers to "the state of lacking the intention to act" because the consequences of the actions have nothing to do with their own but uncontrolled factors (Deci & Ryan, 1985; Vallerand, et al., 1992; Noels, Pelletier, Clément, & Vallerand, 2000). Table 2.8 summarises the continuum of intrinsic motivation, extrinsic motivation and amotivation (Deci & Ryan, 2002). There are three subtypes in intrinsic motivation: to learn; towards achievement; and to experience stimulation. There are four continuous regulation stages in extrinsic motivation: external; introjected; identified; and integrated (Vallerand, et al., 1992; Vallerand, 1997; Deci, Vallerand, Pelletier, & Ryan, 1991). However, even at the last stage of regulation, it is not fully internalised and self-determined (Eccles & Wigfield, 2002).

Table 2.8 Continuum System of Amotivation, EM & IM

Types of motivation	Amotivation	EM				IM
Types of regulation	Non-regulation	External regulation	Introjected regulation	Identified regulation	Integrated regulation	Intrinsic regulation
Extent of internalisation	Non Self-determination----------→Self-determination					

Adapted from Deci & Ryan (2002)

Self-determination Theory can also be applied in language learning. Brown has been one of the main proponents of intrinsic motivation dominant theories (Brown, 1981; 1990; 1994).

He claims that in traditional school settings, extrinsic motivation kills the possibilities of creativity, while intrinsic motivation cultivates a positive, affirming environment (Brown, 1994). However, it is worth pointing out that extrinsic motivation plays a vital role in language learning too, especially when the learners are young, such as in the early years of primary school. In the context of Chinese primary schools, students typically learn English as a subject in schools due to the lack of L2 communities nearby. However, these pupils may be too young to build their own interests in learning English. It is easy for young learners to get bored with language learning. Extrinsic factors, in this case 'Formative Assessment', can potentially help maintain learners' L2 motivation. Self-determination Theory elaborates the dynamic internalisation process of motivation, from non-self-determination to self-determination. It has enriched the studies on motivation and has categorised varied orientations into a systematic framework, which has enabled it to dominate motivation theories for decades.

Synthesis summary of motivation theories

There are seven motivation theories in educational psychology introduced and discussed in this section, which are Expectancy-value Theory, Attribution Theory, Self-efficacy Theory, Achievement Theory, Goal Orientation Theory, Goal Setting Theory, and Self-determination Theory. The interrelationship between each theory is intricate. There are three key elements in motivation theories: individuals, actions, and external environment.

Actions

The variables relating to actions concern task/goal difficulty and specificity (Goal Setting Theory) and its attractiveness (Intrinsic value of Task value in Expectancy-value Theory). These features belong to the actions, known as tasks in Expectancy-value Theory/Attribution Theory or goals in Goal Setting Theory.

Individual: Individual itself

There are two issues with regard to individuals: one is what they think of themselves, which are the ability in Attribution Theory, and self-efficacy as in Self-efficacy Theory. These features belong to individuals, which are separate from the variables relating to actions.

Individual: Individual/Action-Input

The second issue concerning individuals is what individuals think of actions they are going to take, and this is the main focus in most of the motivation theories. There are two types of relationship between individuals and actions: input and output. Input refers to the potential positive effects that action can bring to the individuals. Is the action important to me? (Attainment value in Expectancy-value Theory). If it is not important, then the individuals are amotivated as in Self-determination Theory. If it is important, the following question is 'why is it important?'. Is the action useful to me? (Extrinsic utility value in Expectancy-value Theory, and part of Extrinsic motivation in Self-determination Theory). Is the action incentive to me? (Incentive value of success in Achievement Theory, and to experience stimulation of intrinsic motivation in Self-determination Theory).

Individual: Individual/Action-Output

Output refers to the potential risky considerations (fear of failure in Achievement Theory). This echoes the avoidance state in Goal Orientation Theory (Table 2.7). The cost (in Expectancy-value Theory) includes the efforts and commitment one needs to make (in Attribution Theory and Goal Setting Theory respectively). All the issues considered above concerning individuals and actions, are in the name of Intrinsic motivation (in Self-determination Theory). The purpose is to learn (equivalent to intrinsic motivation in Self-determination Theory, and mastery goal in Goal Orientation Theory).

Environment

Sometimes, the influence comes from beyond the self and action. The last element is the impact that external environments bring to the individuals, such as peer pressure (performance goal in Goal Orientation Theory), extrinsic influences (Achievement Theory and Self-determination Theory), and unstable and uncontrollable factors like luck *etc.* (Attribution Theory). The purpose is to compare oneself to others, by any chance, to best others (Performance goal in Goal Orientation Theory). When taking all the aforementioned elements into consideration, individuals can confirm their probability of success (Achievement Theory) and move towards achievement (Intrinsic motivation in Self-determination Theory).

With the help of Figure 2.7 Construct of Motivation Theories, the map of motivation theories in this study is illustrated clearly. There is no preference for any individual theory, because the main idea of motivation measurement here is to grasp the general motivation, rather than focusing on one particular aspect. Therefore, in the current study, the scale of motivation is established from all the aforementioned theories. However, to map out a clear framework and make it suitable in the context of this study, the theories are arranged for the context of language learning.

2.2.3 Motivation frameworks in language learning

Motivation itself is a complicated concept. However, motivation issues in language learning are even more complex and also unique (Dörnyei, 1998a). Due to the importance of motivation in language learning, a lot of motivation constructs have emerged in second/foreign language studies. In this section, because of their close relationship to the framework used in the present study, the widely-referenced L2 motivation frameworks devised by Dörnyei (1994), Williams and Burden (1997), and Dörnyei & Ottó (1998) will be introduced.

Dörnyei's (1994) Three-level Framework of L2 Motivation

In 1994, Clément, Dörnyei and Noels examined 301 Grade 11 EFL learners with a questionnaire assessing attitude, anxiety and motivation towards learning English, as well as learners' perceptions of classroom atmosphere and cohesion. Teachers participated in the

research by marking the students' proficiency and classroom behaviour, as well as the relative cohesion of class groups.

Based on this study, Dörnyei (1994) built a tripartite motivation construct, including the language level, the learner level and the learning situation level (Table 2.9), which reflect the three aspects of language, including the social dimension, the personal dimension and the educational subject dimension. At the language level, the integrative motives include L2 learners' affective predispositions, *e.g.*, social, cultural, interests in foreign languages. The instrumental motives, which are loosely related to integrative motives, refer to those well-internalised extrinsic motivations. At the learner level, the two components are the need for achievement and self-confidence. They are related to fairly stable personality traits. Underlying the motivational process, several motivational theories are embedded, such as Achievement Theory, Attribution Theory, Self-efficacy Theory and Expectancy-value Theory. At the learning situation level, three areas are discussed, which are the course, the teacher and the group. These areas include teaching and learning environment factors such as the syllabus, the teaching material, the teaching method, the teaching style, the teacher-student relationship, the peer relationship, the learning style, and last but not least, Goal Orientation Theory and Goal Setting Theory. Attempting to summarise a synthesis of various L2 motivation theories, Dörnyei uses this elaborated framework to analyse particular learning situations in depth.

However, according to Dörnyei (1998a), this framework lacks a component about goals and the relationship to Self-determination Theory. Therefore, this model is not perfect. Even though Dörnyei highlights the flaws in this construct, the three-level structure is still a powerful model which echoes the synthesis of motivation theories summarised in the last section of this chapter. It emphasises not only the language and the learners, but also the learning environment. Section

2.2.4 Framework of L2 motivation in this study will describe an L2 motivation framework designed by the researcher which embeds some items that are listed here.

Table 2.9 L2 Motivation Framework by Dörnyei (1994)

Language level	Integrative motivational subsystem Instrumental motivational subsystem
Learner level	Need for achievement Self-confidence • Language use anxiety • Perceived L2 competence • Causal attributions • Self-efficacy
Learning situation level	
Course-specific motivational components	Interest Relevance Expectancy Satisfaction
Teacher-specific motivational components	Affinitive drive

	Authority type Direct socialization of motivation ・ Modelling ・ Task presentation ・ Feedback
Group-specific motivational components	Goal-orientedness Norm & reward system Group cohesion Classroom goal structure

Williams & Burden's (1997) Framework of L2 Motivation

Williams & Burden's extended framework is a list of highly detailed L2 motivation factors (Table 2.10). It categorises all the factors as intrinsic motivation or extrinsic motivation, making up for the aforementioned deficiency in Dörnyei's 1994 framework that Self-determination Theory was not elaborated. Meanwhile, the extended framework introduces personal traits such as age and gender into the list, which brings new perspectives to the current framework of motivation and its analysis later. One of the benefits of this framework is that all the relevant items are listed in one table. However, although the construct is detailed, there are no directional relationships between the items. There is one more drawback to this framework: it is difficult to put into practice in a short interview or in an appropriate questionnaire for young children. So, in a research environment involving young learner participants, the advantage of this framework - that it is detailed - turns into its disadvantage – that it is *too* detailed. Nevertheless, the framework provides a good range of items that can be selected and applied in the framework of motivation in this study, such as 'significant others' (such as parents) in the external factors category.

Table 2.10 L2 Motivation Framework by Williams & Burden (1997)

Internal factors	External factors
Intrinsic interest of activity ・ arousal of curiosity ・ optimal degree of challenge Perceived value of activity ・ personal relevance ・ anticipated value of outcomes ・ intrinsic value attributed to the activity Sense of agency ・ locus of causality ・ locus of control RE process and outcomes ・ ability to set appropriate goals Mastery ・ feeling of competence ・ awareness of developing skills and mastery in a chosen area ・ self-efficacy Self-concept	Significant others ・ parents ・ teachers ・ peers The nature of interaction with significant others ・ mediated learning experiences ・ the nature and amount of feedback ・ rewards ・ the nature and amount of appropriate praise ・ punishments, sanctions The learning environment ・ comfort ・ resources ・ time of day, week, year ・ size of class and school ・ class and school ethos

• realistic awareness of personal strengths and weaknesses in skills required • personal definitions and judgements of success and failure • self-worth concern learned helplessness Attitudes language learning in general • to the target language • to the target language community and culture Other affective states • confidence • anxiety, fear Developmental age and stage Gender	The broader context • wider family networks • the local education system • conflicting interests • cultural norms • societal expectations and attitudes

Dörnyei & Ottó's (1998) Process Model of L2 Motivation

A great advance of Dörnyei & Ottó's framework is that they treat motivation in L2 learning as a dynamic process which contains an action sequence and motivational influence (Figure 2.8). Similar to what Dörnyei has done before, they aim to integrate several studies into a unified framework. This is what I am pursuing in the motivation framework for the current study; therefore, two of Dörnyei's frameworks are studied and used for reference. Although Dörnyei & Ottó have accomplished a well-elaborated model, as with Williams & Burden's framework, it is too detailed to operate under different circumstances in practice.

Figure 2.8 Process Model of L2 Motivation by Dörnyei & Ottó (1998)

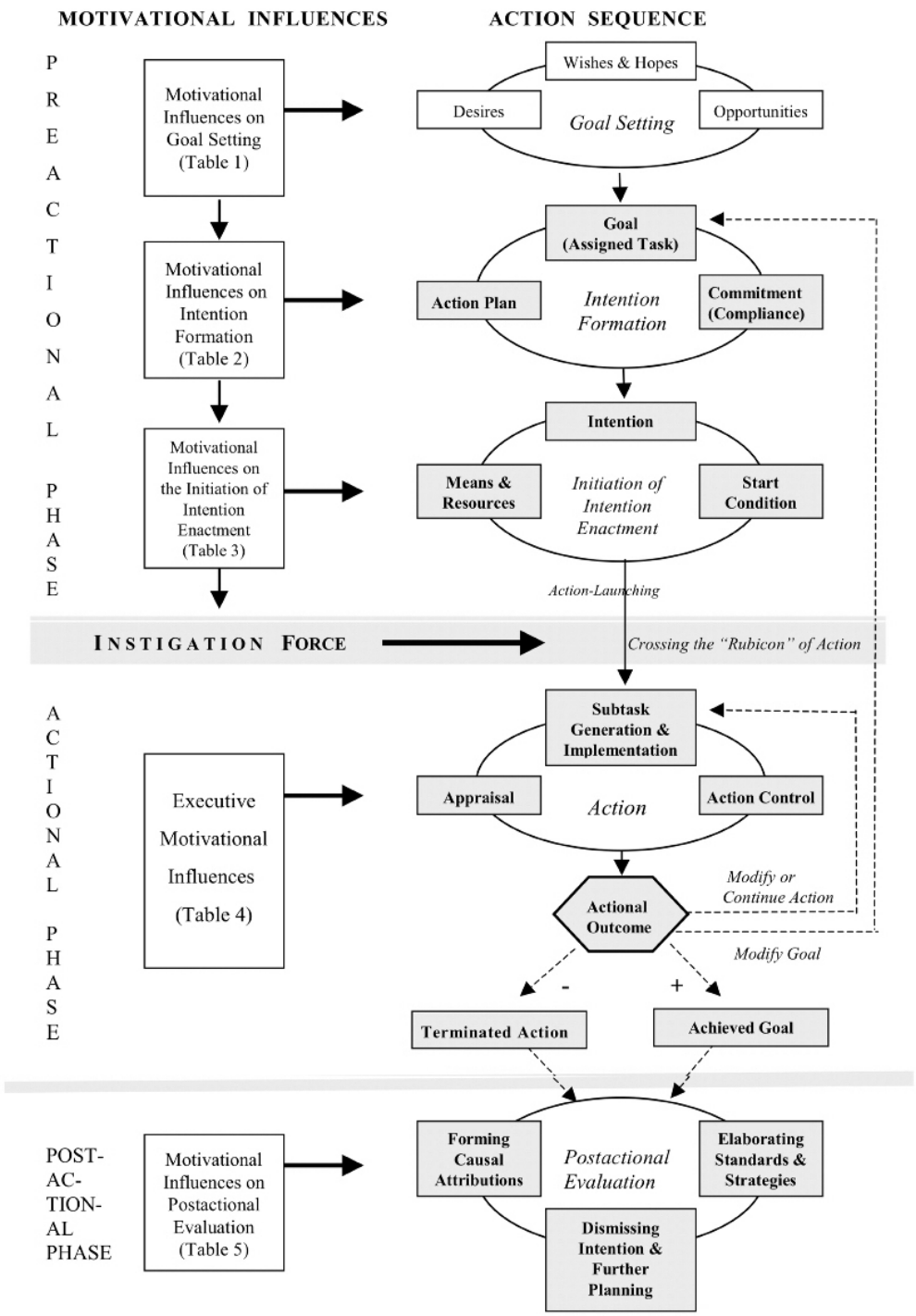

MOTIVATIONAL INFLUENCES
ACTION SEQUENCE
PREACTIONAL PHASE
Motivational Influences on Goal Setting (Table 1)
Motivational Influences on Intention Formation (Table 2)
Motivational Influences on the Initiation of Intention Enactment (Table 3)
Wishes & Hopes
Desires
Goal Setting
Opportunities
Goal (Assigned Task)
Action Plan
Intention Formation
Commitment (Compliance)
Intention
Means & Resources
Initiation of Intention Enactment
Start Condition
Action-Launching
INSTIGATION FORCE
Crossing the "Rubicon" of Action
ACTIONAL PHASE
Executive Motivational Influences (Table 4)
Subtask Generation & Implementation
Appraisal
Action
Action Control
Actional Outcome
Modify or Continue Action
Modify Goal
-
+
Terminated Action
Achieved Goal
POST-ACTIONAL PHASE
Motivational Influences on Postactional Evaluation (Table 5)
Forming Causal Attributions
Postactional Evaluation
Elaborating Standards & Strategies
Dismissing Intention & Further Planning

Table 1. Motivational influences on goal setting

- Subjective values and norms
- Incentive value of goal-related action, outcomes, and consequences (instrumentality)
- Perceived potency of potential goal
- Environmental stimuli; action possibilities; family expectations
- Language/language-learning-related attitudes (integrativeness)

Table 2. Motivational influences on intention formatiom

- Expectancy of success/perceived coping potential
  - self-efficacy/self-confidence
  - perceived goal difficulty
  - amount of expected support
  - L2 anxiety
  - perceived L2 competence
  - L2 contact
  - causal attributions
- Relevance (personal and setting-related); cost-benefit calculations
- Need for achievement and fear of failure
- Degree of self-determination (type of regulation)
- Goal properties
  - goal specificity
  - goal proximity
  - goal harmony/conflict
  - level of aspiration
- Availability of task opportunities and options
- Learner beliefs about L2 learning; knowledge of learning strategies; domain-specific knowledge
- Urgency; external demands; unique opportunity

Table 3. Motivational influences on the initiation of intention enactment

- Action vs. state orientation
- Perceived behavioural control
- Distracting influences and obstacles; number and strength of competing action tendencies
- Perceived consequences for not acting

Table 4. Executive motivational influences

- Selective sensitivity to aspects of the environment
- Quality of internal model of reference
  - action schemata
  - performance standards
- Quality of learning experience
  - novelty
  - pleasantness
  - goal/need significance
  - coping potential
  - self and social image
- Perceived contingent relationship between action and outcome; perceived progress
  - success
  - "flow"
- Sense of self-determination/autonomy
- Teacher's and parents' motivational influence
  - autonomy supporting vs. controlling
  - affiliative motive
  - direct socialisation of motivation
    - modelling
    - task presentation
    - feedback
- Performance appraisal, reward structure, classroom goal structure (competitive. Individualistic, cooperative)
- Influence of learner group (goal-orientedness, cohesiveness, norm and role system, peer role modelling), classroom climate, and school environment
- Task conflict; competing action tendencies; other distracting influences; availability of action alternatives
- Costs involved and natural tendency to lose sight of goal and get bored/tired of the activity
- Knowledge of and skills in using self-regulatory strategies
  - language learning strategies
  - goal setting strategies
  - action maintenance strategies
- Perceived consequences of action abandonment

Table 5. Motivational influences on post-actional evaluation

- Attributional factors: attributional style and biases, prior knowledge about "scripted" events
- Self-concept beliefs: self-confidence/self-efficacy; self-competence; self worth; prior performance history
- Evaluational/attributional cues, feedback
- Action vs. state orientation

The model in Figure 2.8 divides L2 motivation into three phases: preaction, action, and postaction, and each phase has different motivation influences (shown in the following tables). The model has inspired and informed both my FA framework and motivation framework. Firstly, learning with FA is also conceptualised as a process. Therefore, the dynamic movement during the use of FA is emphasised in Figure 2.3 FA framework in the present study. Secondly, the three phases of Dörnyei's motivation framework are applied to the FA framework with a similar action-reaction pattern. Specifically: sharing learning aims in my FA framework, or goal-setting in this framework, is the beginning of all the actions; modifying or continuing actions is actually the core in FA; and starting again with new

learning aims (if the previous ones are achieved) in my FA framework is very similar to elaborating standards and strategies after achieving goals in Dörnyei's framework.

It is not only the FA framework, but also the motivation framework which uses Figure 2.8 as reference. With different learning and teaching activities, learners' motivation changes all the time, hence different motivation theories can be applied. In the preactional phase, Goal Setting Theory stands out with a few further steps such as commitment. As shown in the figure, Goal Setting is not shaded with grey, because Dörnyei believes that it does not reflect reality. It is when students choose to put their beliefs in goals into practice, that the real motivation process starts. By choosing to take an action, the students show their active engagement in setting goals. The model also involves Expectancy-value Theory, Self-efficacy Theory, Achievement Theory, Self-determination Theory in the preactional phase (See Table 1 2 3 in Figure 2.8). In the actional phase, Self-determination Theory and Expectancy-value Theory play a vital role, especially the variables such as extrinsic motives and task values (See Table 4 in Figure 2.8). In the postactional phase, Attribution Theory dominates the influences; this is when students seek to identify the attributions to success and avoid failure (See Table 5 in Figure 2.8). Among all the causal attributions, ability and effort are viewed as the most dominant perceived causes of success or failure in western culture (Dörnyei & Ottó, 1998).

The model combines FA activities and motivation because I believe that FA has at least some influence on motivation, and consequently on achievement too, and it is worthwhile exploring and testing it.

2.2.4 Framework of L2 motivation in this study

After introducing seven motivation theories and three motivation models, it is time to complete the account of the design of a framework that is suitable for the present study. Due to the young age of all the participants, the questions and descriptions are designed to be as simple as possible. Meanwhile, it borrows the 'three-level' hierarchy from Dörnyei (1994) and several items from Williams & Burden (1997). The details of questionnaire design and research methods will be discussed in 3.4.1 Questionnaire. Table 2.11 illustrates the synthesis of varied motivation theories and models in L2 learning context.

Table 2.11 L2 Learning motivation framework in this study

Aspects	Theory	Sub-dimension	Question/Description examples
Language	Value	Attainment value	English is important to me
		Intrinsic value	English is interesting to me
		Extrinsic value	English is useful to me
		Cost	I am willing to offer something (e.g. time) to learn English
	Goal setting	Difficulty (task difficulty)	Learning goals are difficult/easy to me
		Specificity	Learning goals are specific/vague to me
Learner	Goal orientation	Performance goal (to outperform)	I learn English because I want to be a good student or outperform others
		Mastery goal (to learn)	I learn English because I want to master it
	Attribution	Ability	I have the ability to learn English well
		Effort (commitment)	I work hard in English learning
	Self-efficacy	Level	I have high faith in my English
		Strength	I am strongly confident in my English
	Achievement	Need for achievement	I need to succeed in English learning

		(towards achievement) Expectancy of success Incentive value of success (to experience stimulation) Fear of failure	 I want to succeed in English learning I like the feeling of being successful I am afraid of failure
Learning environment	Extrinsic factors	Teacher Parents Peer Education system Culture	I learn because of my teacher I learn English because of my parents I learn English because of friends/classmates I learn English because of school requirements I learn English because I like western culture

In this framework, all the aforementioned motivation theories find their own places and fit into the *language, learner, learning environment* structure. Compared to the Synthesis summary of motivation theories, this framework balances the theories and its variables, known as sub-dimensions in the framework. With regard to *language*, there are two motivation theories, task values in Expectancy-value Theory and Goal Setting Theory. The underlying variables are treated as sub-dimensions to form concrete questions that are about language itself. With regard to *learner*, there are four more theories: Goal Orientation Theory, Attribution Theory, Self-efficacy Theory and Achievement Theory. These provide information about why the learner wants to learn the language. Lastly, Extrinsic factors in Self-determination Theories add the significant others in *learning environment*, which are teachers, parents, peers, education system and culture. Intrinsic factors have already been embedded in other sub-dimensions, so they are not mentioned here again.

2.2.5 Summary

In this section, inspired by Dörnyei and Ushioda's (2013) thorough review of various motivation theories, I borrowed the basic framework of their retrospective review and selected several theories relating to the current research, to illustrate varied motivation theories in a logical sequence. On the basis of the previous explanations in this chapter and the characteristics of classroom FA, I have built a self-designed framework of motivation which can be applied in the current research for young language learners, making motivation researchable and measurable. It is important to depict the information that reflects motivation level in a clear, visual way. In the next section, I will explore the literature on academic achievement and its relationship with FA and motivation.

2.3 Achievement

2.3.1 Defining Achievement

Achievement is a topic that has naturally accompanied the study of learners' motivation in education research (the relationship is explicit in Achievement Theory, for example). However, it cannot be taken for granted in this study. It is natural that education professionals and researchers pay attention to the outcomes of learning. As mentioned in 2.1.9, front-line teachers, who are the ones doing the teaching, prefer reports about exam results to what they perceive as empty theoretical discussion. No matter how influential a research study is, it is not practical if teachers cannot and do not use the results from the

research. To maximise the research impact and its practicability, achievement is the third major theme of this study, recognising that for teachers it is an essential focus.

Although there is a great deal of research on achievement, the construct itself is rarely discussed. For students and teachers in universities, achievement is both an emotional sense of attainment and a quantifiable process or set of outcomes on completion (Jones, 2015). There is no general agreement on how achievement is best evaluated, or which aspects are the most important. Thus, a useful definition of achievement is a priority in this study. Policy-makers in the education system claim that schools have to fulfil society's mandate. In other words, schools need to help students to accomplish academic achievements. Academic achievement usually is considered to be a multifaceted construct that encompasses the domain of learning. Spinath (2012) stresses that different perspectives can be taken into consideration in achievement, such as the comparison between individuals and societies, as well as psychological and educational views. Yet, most of the achievement indicators refer to knowledge acquired in an educational system, curricula-based achievement tests, and educational degrees or certificates (Steinmayr, Meißner, Weidinger, & Wirthwein, 2017).

Every curriculum subject need to be contributed to such achievements; hence, academic achievement sometimes refers to students' overall achievement in schools (Murray, Low, Hollis, Cross, & Davis, 2007). Of course, achievements in individual subjects are often discussed separately in research. For instance, in Zach, Shoval and Lidor's review on physical education and academic achievement (2017), the term *academic achievement* was searched for along with some substitutable key terms such as *academic performance*, *school achievements* and *testing results*. The different operational definitions have led to inconsistent findings in the literature.

Some of the work regards educational issues as academic achievement variables, such as grit, educational purpose *etc*. (Seabi & Payne, 2013; Tucker-Drob & Briley, 2016). These relate to the internal progressive contours of learning, such as improving, mastering and overcoming, which emphasise the effort and perseverance. Other work has tended to categorise aspects of achievement as external quantifiable elements for measurement. Some of them use a very complex equation called the 'value-added model' (Rivkin, Hanushek & Kain, 2005; Rochoff, 2004; Hanushek, Kain, Markman & Rivkin, 2003). However, most researchers see achievement as relating to tests or high-stake examinations, from global indicators like school grade point average (GPA) to specific indicators like standardised test scores (Fan & Chen, 2001).

Short-term research can provide a snapshot of learners' achievement, but such a picture of their outcomes is relatively incomplete and not fully accurate. Because of the short time length of this study, it was important to collect more information to enrich the resulting picture as much as possible. Thus, in this study, I attempt to capture achievement in language learning across skills including listening, speaking, writing and reading. The next sections will clarify the difference between achievement, attainment and progress; address the special status of achievement in China; narrow down the conceptual focus of achievement in language learning; and report the significant gender differences in

achievement. Following this, the conceptual and operational definitions of achievement will be stated.

2.3.2 Achievement *vs.* attainment/progress

The terms achievement and attainment are often used interchangeably in many studies. However, they are different. What I am pursuing in this study is achievement, rather than attainment or progress.

Attainment has clear benchmarks that can differentiate lower and higher achievers. It shows clearly which level students are on. For example, in China, the normal scale of test is 100 points, and 60 is the benchmark for a 'fail' or a 'pass'. Students who attain the benchmark or standards are considered as average students; those who do not attain the benchmark are lower achievers. The attainment figures show the performance of a student at the end of a period of time. This index is important because pupils have varying levels of ability and starting points, and it gives clear and straightforward results of that; furthermore, the results are associated with a range of outcomes, such as income, employment and health, which can be used as predictors for the future (Myatt, 2013).

Progress refers to how well pupils have advanced in their learning from their starting points and capabilities. Some pupils may seem to be low attaining, but the progress they have made may be greater than other pupils starting from a higher baseline. Therefore, it is unfair for them to be considered as lower achievers only, because it ignores their efforts in making progress.

Achievement comprises both attainment and progress. In OFSTED terms, it is the distance travelled between two points in time (Whetstone, 2011). It refers to outcomes now in comparison to what a student had in the last exam, or the extent to which a student has accomplished specific goals in instructional environments, which is a relative concept that measures improvement of a student (Steinmayr, Meißner, Weidinger, & Wirthwein, 2017). Table 2.12 presents the comparison of attainment, progress and achievement. It distinguishes between the terms.

Table 2.12 Attainment, progress and achievement

Term	Description
Attainment	This is the standard of academic attainment, typically shown by test and examination results
Progress	This is the extent to which pupils have progressed in their learning from their starting points and capabilities
Achievement	This considers the standards of attainment reached by pupils and the progress they have made to reach those standards

To sum up, achievement is the progress made by students in acquiring new knowledge or skills, which considers starting points and improvement; progress is the actual improvement; while attainment is reaching a certain level that has been set as a benchmark.

2.3.3 Achievement in China

In Chinese society, the accomplishing of academic achievements is one of the most important tasks for school-age children (Fu, Chen, Wang, & Yang, 2016). The nature of the importance of achievement in Confucian culture may differ from that in other western countries.

First of all, achievement is important to students themselves. As a higher-achieving student, one is more easily accepted by peers and seniors, such as teachers and parents. In contrast, the lower-achieving students are often regarded as problematic (Phillipson & Phillipson, 2007). Secondly, students' achievement is not only important to children, but also their parents and family, because failure in academic achievement may bring disgrace and shame to the parents, even the ancestors (Ho, 1986). As the only child in a 4-2-1 family structure (this will be elaborated in 3.8.1 Doing research in primary schools in China), they receive more attention than children with siblings in previous generations received. More responsibilities are put on the child's shoulders. Lastly, students' academic performance sometimes determines teachers' wages (along with many other factors, such as number of years teaching, the professional title, *etc.*). Thus, teachers' dominant focus may be on students' achievement even if that dominance is unconscious. If there is no relation between teachers' wages and students' achievement, the teachers may be able to focus more on students' learning process, learning skills and learning autonomy. However, economic pressures tend to force teachers to emphasise achievement, rather than the aforementioned learning dimensions.

Therefore, Chinese children's academic achievement is closely related to everyone around them. Students may endure much more pressure than children and young people in western cultures in terms of the drive for higher achievement, because it is not only about themselves, but also about their family and teachers.

2.3.4 Achievement in language learning

Many indicators are applied in the measurement or collection of data on achievement in language learning. This section gives several examples of research on English language learning achievement using a range of indicators.

Some of the examples are indicators with specific criteria. In a language anxiety study, Yan and Horwitz (2008) list the range of achievements in English: doing well in a test, speaking with accent-free pronunciation, speaking fluently, communicating with native speakers in the target language, functioning well in normal class activities.

Moreover, many so-called 'achievements' take the form of different types of tests. Thomas and Collier (2002) defined English learners' long-term achievement on a national standardised test in English Total Reading, which is a subtest measuring academic problem-solving across the curriculum. English Total Reading results are the most closely correlated to the standardised test required for admission to a higher level of study. Similarly, the GCSE (General Certificate of Secondary Education) grades of English language or English literature are used for English achievement (Rimfeld, Kovas, Dale, & Plomin, 2016). GCSE is a nation-wide exam that students can take at the end of compulsory schooling, usually at age 16. A

reading comprehension test is used for reading achievement from Woodcock Johnson Tests of Achievement-III (Woodcock, McGrew, & Mather, 2001).

Therefore, in language learning, tests are generally used to measure achievement. Because of that, this study will use a similar form of test-quiz to assess achievement in English learning. Although there are drawbacks to the quiz (which will be discussed in 3.9 Reflection and reflexivity-Achievement quiz), it is judged to be the most convenient and valid tool to measure achievement in this study.

2.3.5 Achievement and motivation

The relationship between achievement and motivation is easy to take for granted. Plenty of studies have explored the relationship between their research target and achievement/motivation respectively, such as classroom belonging (Goodenow, 1993), social competence (Zsolnai, 2002), or the 'big five' personality traits (Komarraju, Karau, & Schmeck., 2009). It is true that many researchers have proved a close relationship between achievement and motivation (Pintrich & De Groot, 1990; Broussard & Garrison, 2004). Self-efficacy, goal-setting, perceived academic competence, self-determination, and prior grades can positively influence and even predict students' academic achievement across subjects like mathematics, combined science *etc*. (Zimmerman, Bandura, & Martinez-Pons, 1992; Covington, 2000; Fortier, Vallerand, & Guay, 1995; Akomolafe, Ogunmakin, & Fasooto, 2013; Chow & Yong, 2013). Among many motivational attributions, goal orientation was not a significant predictor of achievement (Kitsantas, Steen, & Huie, 2017). However, there are other attributions, such as talent, luck and task difficulty, that would affect achievement even if a learner is highly motivated (Weiner, 1992). Hard work is the essential element in academic success and there has been widespread recognition of this (Smith & Skrbiš, 2017). For Asians, parents and peers all hold high standards, and young learners study diligently because they believe that the road to success is through effort (Chen & Stevenson, 1995). Therefore, it is possible to find situations where students are motivated but not accomplished. Out of an interest in that phenomenon, the researcher wishes to explore the relationship between participants' achievement and motivation in this study.

2.3.6 Achievement between genders

Achievement has been proven to be both an antecedent and consequence of several emotional, behavioural and social characteristics (Pop, Negru-Subtirica, Crocetti, Opre, & Meeus, 2016). Gender is one of them. The differences in achievement and its connection to other differences between boys and girls is significant and has been observed in the literature of many fields. For example, the bidirectional association between depressive problems and poor academic achievement is only found in girls (Verboom, Sijtsema, Verhulst, Penninx, & Ormel, 2014). Peer rejection and discrimination are associated with lower achievement particularly among girls (Bakker, Denessen, Bosman, Krijger, & Bouts, 2007). High academic achievement and depressive symptoms are only negatively related in boys (Fröjd, et al., 2008). Boys show lower levels of motivation, engagement and achievement due to the negative attitudes of their peers toward school (King, 2016). The findings suggest that students' achievement is related to their interpersonal relationships and show a significant difference between genders. Interpersonal relationships are

important to assessment, particularly to student-led assessment because the young children seem more emotional than other age groups. Boys and girls with varied achievement levels show different ways of treating others, which may affect their preference in using FA.

When it comes to achievement on its own, girls have been shown to outperform boys across countries and subjects (King, 2016; Voyer & Voyer, 2014). In the UK and Australia, girls have consistently outperformed boys in standardised examinations such as GCSE since the 1990s (Foster, 2000). The explanations for this could be the unique social and cultural contexts of the UK, including equal opportunities in schooling and rates of male unemployment (Weiner, Arnot, & David, 1997). In Asia, similar results are found. In the school subjects English and Chinese, a greater proportion of girls achieve excellence, but in mathematics boys tend to do better (Wong, Lam, & Ho, 2002; King, 2016).

To sum up, significant gender differences in achievement have been found by researchers, especially in language learning. However, the differences between boys and girls in correlations of achievement and other factors are inconsistent and need to be explored individually.

2.3.7 Achievement measurement

The measurement of achievement can be approached with different statistical models, such as average points scores, value-added and contextual value-added models (Whetstone, 2011). As mentioned earlier, several researchers use a general achievement equation which describes achievement as a value-added model. They link the achievement model with a few variables that fit their studies, such as students' fixed characteristics, teachers' effect, peers' effect, school-year effect, family effect, and random error, *etc.* The equation below presents an example of how it works in the educational situation (Hanushek, Kain, & Rivkin, 1998).

$$\Delta A_{igst} = SP_{igst}\lambda + X_i\beta + D_{gst}\theta + C_s\eta + \gamma_i + \delta_{gst} + \varepsilon_{igst}$$

In this equation, the test score (A_{igst}) stands for students *i* in grade *g* in school *s* in year *t*. The variables are: special education status (SP), vectors of family characteristics (X), school demographic characteristics (D), and community type dummy variables (C); and the three error components are: a time invariant individual component (γ_i), a school quality component that varies across grades (δ_{gst}) and a random error (ε_{igst}).

Compared to the value-added model, the contextual value-added model takes additional factors into consideration, for example, ethnicity, gender, and the Income Deprivation Affecting Children Index, *etc.* (Whetstone, 2011). However, it seems these models are too complex for application to the present study, given the research size, scale and focus.

On one hand, the measurement of academic achievement sometimes does not relate to actual scores directly or indirectly. There are some studies focusing on the traits featured in academic achievement. For instance, grit, intellectual curiosity, intellectual self-concept, mastery orientation, educational value, intelligence mind-set, and test motivation are considered in Tucker-Drob and Briley's work (2016) about intervention-oriented and/or social theories. Seabi and Payne (2013) regard educational purpose, academic autonomy,

expectations about academic success *etc.* as academic achievement variables. In Machell *et al.*'s work (2016), students are asked to complete face-valid items about their academic performance. It mixes students' actual grades and their academic places in the class comparing to peers in their own minds. However, this type of measurement focuses on children's mental development, rather than the improvement of academic performance. Therefore, this type of measurement is not used in this study.

On the other hand, achievement measurement could relate to actual stores. Despite the criticism of its flaws, such as grade inflation, different grading criteria *etc.*, GPA has been the most often used and most reliable indicator for achievement until now (Pop, Negru-Subtirica, Crocetti, Opre, & Meeus, 2016). Other researchers who study the relationship between FA and achievement also use summative tests as the way of measuring achievement (Carrillo-de-la-Peña, et al., 2009; Sheard, Chambers, & Elliott, 2012). The current study examines achievement over the short term (approximately three months). Thus, dynamic summative scores will play a significant part in measurement of achievement in English learning.

There are many types of achievement. Fifteen types are identified and coded in a review of the relationship between self and achievement by Hansford and Hattie (1982) and these include I.Q., work study and GPA *etc*. Five of them are related to verbal facility or ability, which are reading, vocabulary, verbal skills, language use, English comprehension. There is no significant difference between each achievement type, statistically. As mentioned in Yan and Horwitz's research (2008), four main skills (listening, writing, reading, speaking) in English learning are important. It is difficult to measure four skills without a concrete instrument. Therefore, the researcher chose to use a short, quick quiz after each unit in the English textbook to quantify the general achievement of students. The use of language skills is embedded in self-designed quizzes. The scores that students achieve in respect of each skill make up one final score which gives a snapshot of a student's overall achievement. The dynamic summative scores from these quizzes during the fieldwork represent the learners' general achievement in English.

2.3.8 Summary

The concept of achievement and its measurement were reviewed in this section. The definition of achievement was discussed because in many studies it is used interchangeably with other terms such as attainment. I have also explained the special status of 'achievement' in China due to the pressure from oneself, one's family and one's teachers. To relate the literature review to this study, achievement in language learning has been addressed with several examples of previous studies. Solid rationales were provided for choosing elements to be used in the models and methods for the current study. In relation to gender, it was noted that boys and girls show significant differences in achievement. These may be the consequences of innate differences or social factors; although that wider debate is beyond the scope of the present thesis, social or environmental influences potentially include the topic being explored here: formative assessment. More discussion on this area will be presented in the Discussion chapter. Lastly, the type and indicators of achievement for the purposes of this study were described. A language quiz was identified

as a tool to quantify achievement, as this could yield dynamic results week to week across the fieldwork period.

The literature on formative assessment was reviewed. The impact of formative assessment in the classroom is a subject of debate in terms of whether it has positive effects on motivation and achievement. Therefore, the aim of this study is to explore the relationship between FA and motivation/achievement and the impact of FA on these variables.

This chapter has mapped out a clear theoretical framework for the study, having reviewed the relevant literature on the three main themes: formative assessment, motivation and achievement.

As the Chinese idiom says, “to do a good job, an artisan needs the best tools”. In the next chapter, methodology and methods chosen for the study will be described and justified.

Chapter 3 Methodology and Methods

3.1 Introduction

Many new researchers are uncertain of the distinction between *methodology* and *methods*. Fortunately, there is a metaphor which explains the difference well: methods are the ingredients of research and methodology provides the reasons for using a particular research recipe (Clough & Nutbrown, 2007). The methodology is the *procedure design* and the methods are the *tools* that we use to acquire knowledge. With these principles in mind, this chapter discusses the 'what', 'why' and 'how' for decisions made and actions taken regarding the methodology and methods in this study.

3.2 Research aims and research questions

The first consideration for a proposed study is why it should be done. Extending the metaphor above, that is - why should the dish be cooked? The current study sought to answer a core question: in English language learning, what is the impact of FA on primary school students' motivation and achievement? To answer the core question, the study set two aims: to identify the relationships between FA strategies and students' motivation/achievement; and to understand the impact of FA strategies on students' motivation/achievement; from different stakeholders' points of view. Having articulated the overall aim of the study, the next step was to specify the research questions that would need to be answered in order to satisfy the ultimate aim of the research (Moore, 2006).

3.2.1 Research questions

According to the research aim, the initial question was: is there an impact of FA on motivation or achievement? Therefore, the relationships between these aspects were to be investigated by three specific questions of correlation:

1) Is there a relationship between FA and students' motivation?
2) Is there a relationship between FA and students' achievement?
3) Is there a relationship between students' motivation and achievement?

In light of contexts in the field, the difference in the relationships was investigated in three dimensions: achievement level, FA environment and gender. Therefore, the first three research questions were extended by three further sub-questions:

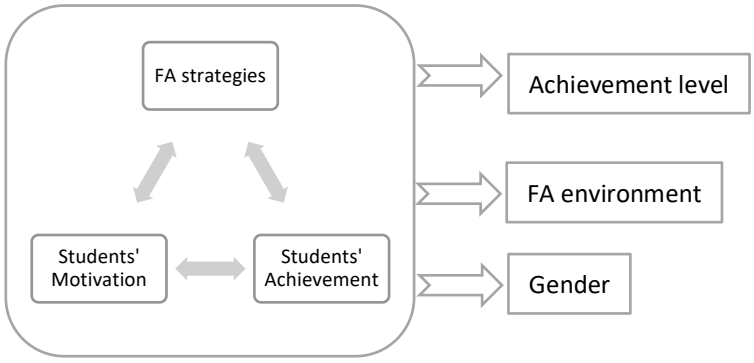
1) Is there a relationship between **FA** strategies and students' **motivation**? (4.6.1)
    a. Is there a difference in the relationship between FA and motivation in different **achievement groups**? (4.6.2)
    b. Is there a difference in the relationship between FA and motivation in different **FA environments**? (4.6.3)
    c. Is there a difference in the relationship between FA and motivation in different **genders**? (4.6.4)

2) Is there a relationship between **FA** strategies and students' **achievement**? (4.7.1)
   a. Is there a difference in the relationship between FA and achievement in different **achievement groups**? (4.7.2)
   b. Is there a difference in the relationship between FA and achievement in different **FA environments**? (4.7.3)
   c. Is there a difference in the relationship between FA and achievement in different **genders**? (4.7.4)

3) Is there a relationship between students' **motivation** and **achievement**? (4.8.1)
   a. Is there a difference in the relationship between motivation and achievement in different **achievement groups**? (4.8.2)
   b. Is there a difference in the relationship between motivation and achievement in different **FA environments**? (4.8.3)
   c. Is there a difference in the relationship between motivation and achievement in different **genders**? (4.8.4)

To summarise diagrammatically, Figure 3.1 shows that the quantitative research questions were embedded in the relationships between FA, motivation and achievement, in three contexts: different achievement levels, different FA environments and different genders.

Figure 3.1 Research framework – quantitative



Aside from the questions of whether positive or negative, strong or weak relationships would emerge between the above factors, another two dimensions were explored: English teachers' and English learners' perspectives on FA. The questions specified were as follows:

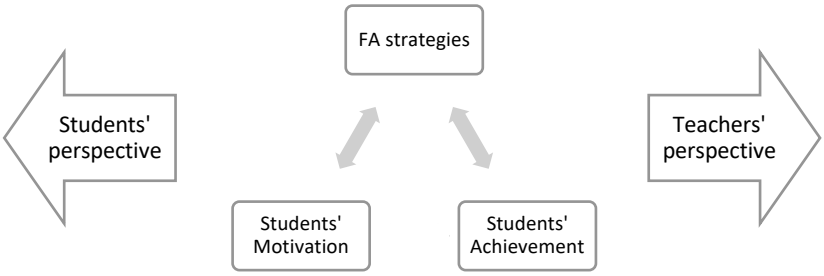
4) What is the impact of **FA** from English **students'** perspective? (5.2)
   a. What is the impact of FA strategies on **motivation** from English **students'** perspective?
   b. What is the impact of FA strategies on **achievement** from English **students'** perspective?

5) What is the impact of **FA** from English **teachers'** perspective? (5.3)
   a. What is the impact of FA strategies on **motivation** from English **teachers'** perspective?

b. What is the impact of FA strategies on **achievement** from English **teachers'** perspective?

depicts how additional qualitative information and insights could be obtained from learners' and teachers' views.

Figure 3.2 Research framework – qualitative



3.2.2 Operational definitions

Operational definitions help to focus the research questions and decide what data to gather to address those questions. There were three topics in this study: formative assessment, motivation and achievement; two main participation groups: students and teachers. According to the Chapter 2 Literature review in the current study, these terms were defined as:

Formative assessment: (2.1.1) a type of assessment used by any stakeholder in the classroom, for the purpose of collecting information and adjusting teaching delivery and learning accordingly, to help students learn better. It comprises mainly five strategies: sharing learning aims, questioning, feedback, peer-assessment and self-assessment. (Other more minor formative assessment strategies have been identified, but these are not discussed in the present study.)

Motivation: (2.2.1) the beliefs, values and goals that determine which task students pursue and their persistence in achieving them, and the standards they set to determine when a task has been accomplished. This was measured by the synthesis of students' responses to 20 motivation questions which were drawn from varied motivation theories and models in L2 learning contexts.

Achievement: (2.3.1) students' comprehensive and dynamic academic performance in English across the four skills: listening, speaking, writing and reading, including their starting point and progress. This was measured by English quizzes during the 10-week fieldwork period.

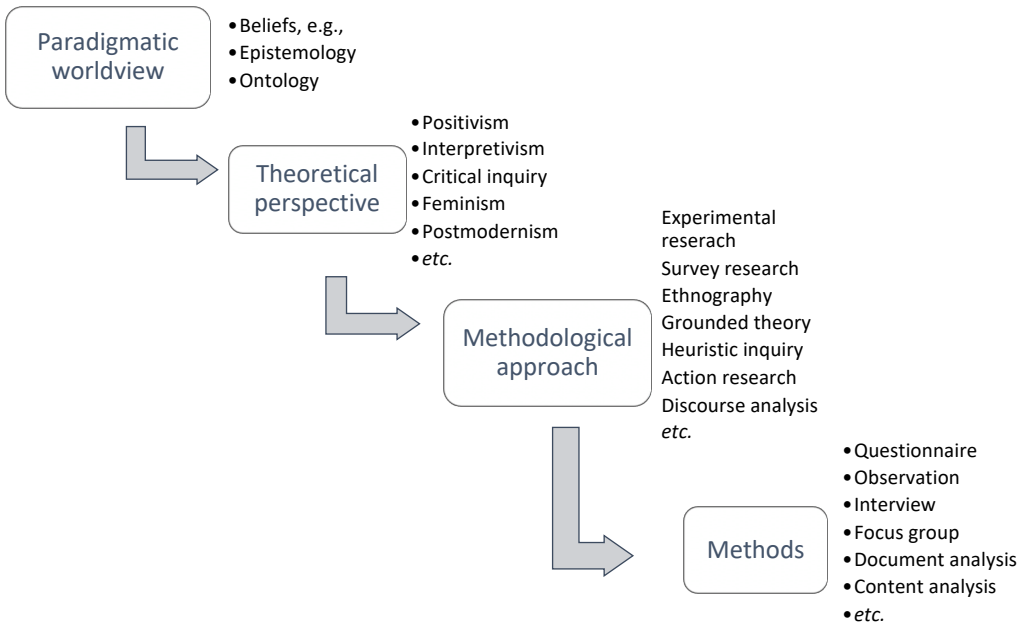
Students: Chinese primary school students who were learning English in Jiangsu Province.

Teachers: Chinese primary school teachers who were teaching English in Jiangsu Province.

3.3 Philosophical foundations

Research studies done by/with people are not always objective. There may be biases at many stages in the research process, from formulating research questions to consideration of findings, due to different theoretical standpoints and beliefs. Philosophical considerations are always the first step to understanding a piece of research. Ontology is closely entwined with epistemology at the top of a research design hierarchy (Figure 3.3). Methodology, then, takes a place in the middle, while method is the last consideration (Hammond & Wellington, 2013).

Figure 3.3 Research design hierarchy



Adapted from Crotty (1998).

3.3.1 Ontology and epistemology

Ontology - pragmatism

Ontology is the nature of reality, in particular social reality. It embodies *what is reality*, while epistemology refers to the *nature* of knowledge, seeking to understand *what we know* and *how we know* (Gray, 2014; Neuman, 2014; Crotty, 1998). Both ontology and epistemology inform theoretical perspectives in research, to answer the question *what approach can be used to get knowledge*. Below, ontology and epistemology are discussed in more depth in relation to the current research study

Ontology is about how people see the world. The world is conceptualised as a set of objective entities that pre-exist, waiting for people to discover them, or as a dynamic process that is shaped by humans' interpretations (Neuman, 2014). Accordingly, there are two main opposing, mutually exclusive positions in ontology: objectivism and subjectivism.

The first standpoint is that there is “an objective reality which exists independent of the observer”. The second standpoint understands reality as “it appears subjectively or, more commonly, as negotiated within groups” (Hammond & Wellington, 2013).

The facts of the world are not just the concrete entities - take this study as an example. Whereas researchers may see formative assessment as a positive action in teaching, some front-line teachers may understand it and treat it merely as a compulsory school requirement or pedagogic implementation and one that generates too much extra workload. One person's trash is another's treasure, but they actually are referring to the same ‘thing’. As for the *objective reality*, can people ever know it? The answer is doubtful. Humans will know the *truths* or the *facts* of a given phenomenon better and better, but only marginally closer and closer to the *objective reality*.

Therefore, the ontological position adopted in this study was pragmatism, which seeks to reconcile subjectivism and objectivism (Johnson, Onwuegbuzie, & Turner, 2007). It suggests employing the most useful approach to answer research questions, rather than engaging in an ‘egotistical’ debate over qualitative or quantitative approaches (Cohen, Manion, & Morrison, 2011). There is a useful metaphor here. Conceptualising the research is like a gift packaging service: rather than putting the gift into a tall box named objectivism or a short box named subjectivism, the gift, or the research in this case, will be wrapped according to its own shape named pragmatism.

Epistemology - positivism and interpretivism

Epistemology refers to the understanding of how people know the world and what makes a claim truthful (Neuman, 2014). A dichotomy of epistemology stances is presented in social science (Gray, 2014). Objectivism believes that truth exists apart from the operation of any consciousness, and this correlates to positivism in its theoretical perspective; constructivism, on the other hand, claims that truth is imposed on the object by the subject’s interactions with the world, that human beings and the world have influence on each other, which correlates to interpretivism in its theoretical perspective.

An ontological stance is closely related to a particular epistemological stance and *vice versa.* The consequences of choosing pragmatism as the ontological stance are that the research is driven by the research questions rather than the preferences of the researcher. The research questions require both quantitative and qualitative data to answer them; accordingly, the epistemological positions and theoretical perspectives adopted in this study are positivism and interpretivism (Crotty, 1998; Cohen, Manion, & Morrison, 2011).

Positivism implies a stance where the social scientist is an observer of social reality (Cohen, Manion, & Morrison, 2011). The properties of the social world can be scrutinised and measured directly via observation (Gray, 2014). Following the principle of deductivism, the hypotheses of relationships between FA, motivation and achievement can be tested and defined; and following the principle of inductivism, knowledge about these three topics is provided by the gathering of the facts (Bryman, 2016).

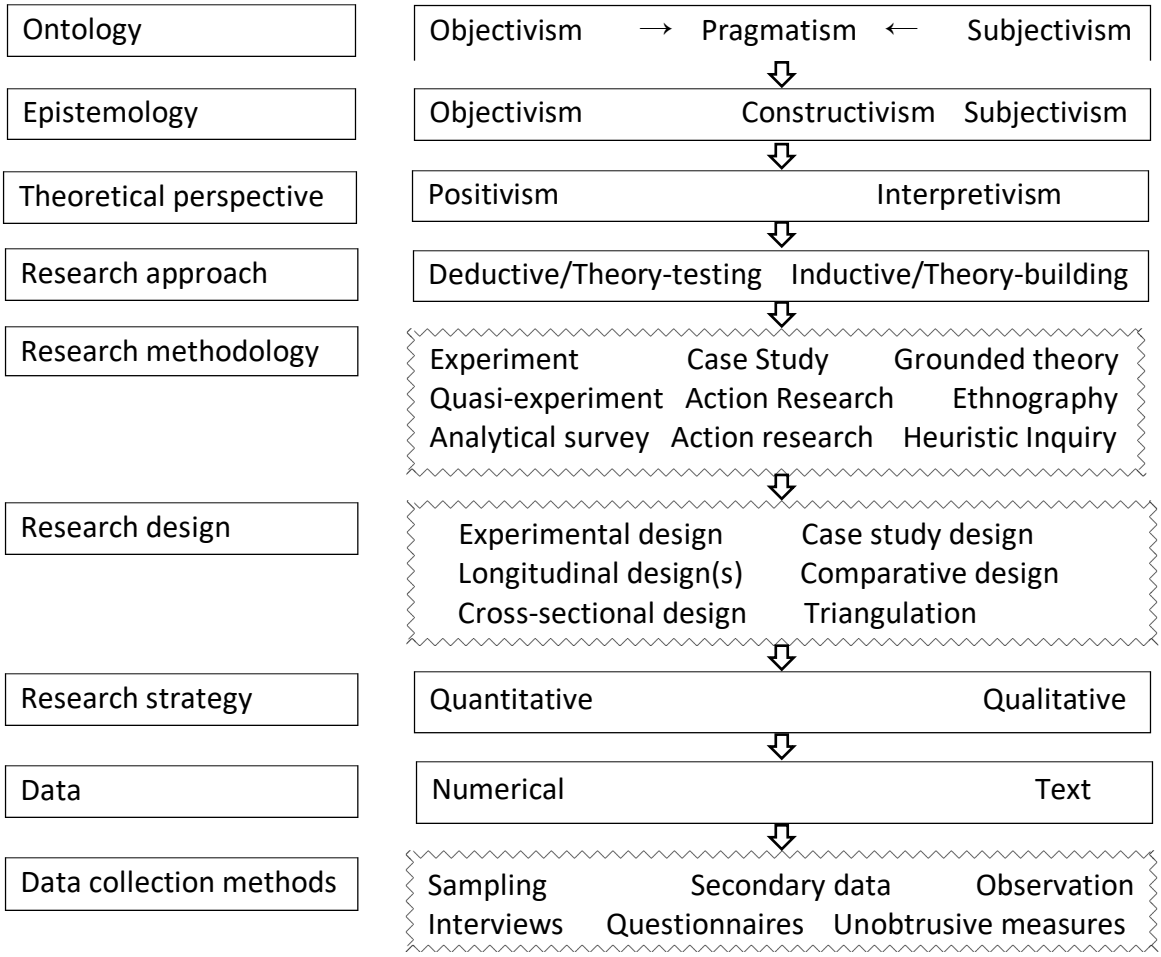
Interpretivism here requires the social scientist to understand the subjective meaning of social action, which respects the differences between people (Bryman, 2016). The existing relationships between FA and motivation/achievement are merely the *facts* that can be measured. This study pursues the *reasons* behind the relationship too. In other words, to know the impact of FA on motivation and achievement, it is necessary to know not only about the *facts* of the relationships, but also the *hidden explanation* for the relationships.

3.3.2 Methodology and methods

Methodology – Analytical survey and Phenomenological research (semi-structured)

The 'recipes' used in this study were: (a) analytical survey, which attempts to test a theory by exploring the relationship between variables; and (b) phenomenological research, which picks up people's experiences or perspectives that were not captured by the survey as part of the relationship between the variables (Gray, 2014). The methodologies were generated from the quantitative and qualitative dimensions of the research question respectively, which combined the deductive approach and inductive logic (Figure 3.4).

Figure 3.4 The elements of the research process

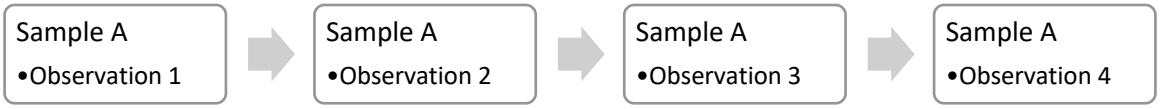


Adapted from Sauders, Lewis, & Thornhill (2000); Cohen, Manion, & Morrison (2011); Bryman (2016)

Research design – Cohort study and Comparative study

This study was designed as a combination of cohort study and comparative study (Figure 3.4). A cohort study is a form of longitudinal design. A cohort is a group of people who share a defining characteristic (age, learning environment, textbook and FA environment in this study). The research model for the cohort study is shown in Figure 3.5. Cohort studies are particularly appropriate for the attempt to link the factors and outcomes when the dependent variable changes (Cohen, Manion, & Morrison, 2011). However, the dependent variable in this study – the use of FA – cannot be changed, due to the ethics and exigencies of the situation. Hence, two FA environments were compared in order to explore the impact of different uses of FA. Comparative study allows us to understand social phenomena better when the contrasting situations are compared (Bryman, 2016).

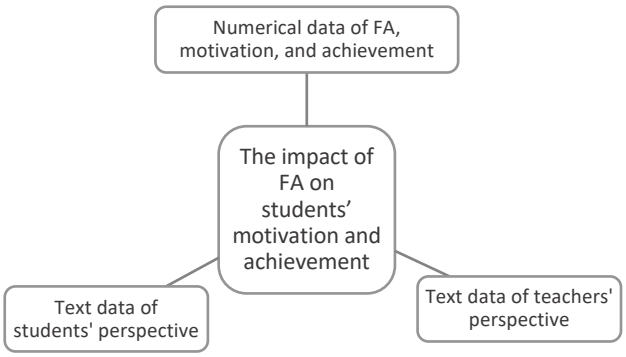
Figure 3.5 Model of Cohort Study



Triangulation, which is the combination of methodologies in the study of same phenomena (Denzin, 2006), uses two or more independent measurement processes to gather data, and explicitly requires comparison of different sets of data, which can compensate for potential weaknesses in each data collection method and greatly reduce the uncertainty in interpreting data (Hammond & Wellington, 2013; Webb, Campbell, Schuwartz, & Sechrest, 2000; Easterby-Smith, Thorpe, & Lowe, 2002; Flick, 2014; Jick, 2006).

The focus of the research was to explore the impact of FA on students' motivation and achievement. Results are triangulated using three sources (Figure 3.6): the numerical data for FA, motivation and achievement, which can be used to test the relationships; the textual data from students' and teachers' perspectives of the impact, which reveal the impact directly.

Figure 3.6 Triangulation of the research design



Methods – questionnaire, quiz and interview

Research methods are not determined dogmatically by assumptions that flow from one paradigm or another, but naturally flow from the research questions. Appropriate research methods offer the best chance of obtaining useful and workable answers to the questions

(Gray, 2014). Therefore, the study selected particular 'ingredient' methods: structured questionnaires and semi-structured interviews, to explore the quantitative and qualitative questions respectively. To be more specific, a motivation questionnaire was used to obtain information about students' motivation; an FA questionnaire and an interview were used to investigate the use of FA and students' perspectives on FA; an interview was used to obtain teachers' perspectives on FA; a language quiz was used weekly to record the students' achievement.

Questionnaires usually consist of a list of statements or questions. Respondents indicate their preferences by, for example, ranking levels of agreement, to generate numerical data from the questionnaire. There were two reasons for using the questionnaire method. Firstly: efficiency. Methods involving time-consuming interpersonal interactions, such as individual interviews or case studies, are not always feasible if the research seeks to obtain information from hundreds of respondents. A self-administered questionnaire, once designed, piloted and modified, can be applied in the study with an ideal number of respondents, no matter how large or how small that optimal number is. Secondly, the absence of interviewer effects is an important concern. When respondents complete questionnaires, they are away from the influence of interviewers. They will not feel the need to be appealing to interviewers if they perceive an interviewer as someone important or if they want to impress them. Similarly, respondent behaviours will not elicit bias in the interviewers.

An English language quiz designed and produced by the researcher was used to measure the students' attainment changes. There were several types of tests in the schools during an academic term: monthly test, mid-term test, final test, and unified test. The first three types of tests were school-based and the last one was city-wide. These tests were norm-referenced tests. They compare a student's achievement to other students' achievement (Cohen, Manion, & Morrison, 2011). However, the fieldwork commenced at the beginning of a term, when the students had not yet completed any of their usual tests. The researcher-produced quiz, a criterion-referenced test, was therefore the best option to obtain records of achievement in this study. It provided the information about exactly what a student has learned (Cohen, Manion, & Morrison, 2011).

However, the questionnaire and quiz were by themselves not sufficient to obtain all the information needed to answer the research questions. One of the biggest problems with a questionnaire is the depth of question. Open-ended questions were not used, given the age group of the children (10 years old) and considering the typical ability of children this age to organise their answers in a short time. Closed questions limit respondents' ability to express themselves and do not enable collection of additional data. Therefore, semi-structured individual interviews with 12 students and two teachers were used to supplement the quantitative data, to give a much richer picture for this study.

The research interview, to cite Gray (2014), is the most logical research technique if the research aims to examine feelings or attitudes. Wellington and Szczerbinski (2007) also point out that interviews can explore interviewees' thoughts, values, prejudices, perceptions, views, feelings and perspectives. The current study is pursuing knowledge not only of the relationship between FA and motivation/achievement, but also the views and

thoughts on the phenomenon from teachers' and students' perspectives. Quantitative data can identify a relationship; the qualitative task of exploring that relationship rests on interviews. According to the degree of formality required, there are three types of interviews: unstructured, semi-structured and structured. The semi-structured interview was chosen because pre-set topics and questions can help guarantee the basic information while leaving some spaces for interviewees to express themselves freely. Moreover, semi-structured interviews allow interviewers not only to measure what is intended to be measured, but also to explore issues in more depth in a comparatively short space of time.

This section discussed the justification for methodologies and methods applied in the current study. In the next section, the design of research instruments will be introduced.

3.4 Research instruments

Three research instruments were used. These were the structured questionnaire (motivation and FA), the English quiz and the semi-structured interview.

3.4.1 Questionnaire

The questionnaire is attached in Appendix 2 (and its English translation as Appendix 2.1). There are five parts to the questionnaire: introduction, instruction, questions, personal information collection and researcher contacts.

The document begins with an introduction to the researcher's background then sets out: the aim of the research; the structure of the questionnaire; the definition of the key concept, formative assessment; the time length it will take; the use of findings; the outputs; and the research ethics. This information was followed by instructions about how to respond to the questions. At the end of the questionnaire, researcher contact details are shared, and information about students' registration number and class were collected. The researcher could identify the respondents' gender, achievements and their FA environment, with the information provided.

The main part is the questions. There are two sets of questions: motivation and FA. The first set of questions concerning students' motivation in language-learning was adapted from the motivation framework (Table 2.11). Questions took the form of simple statements, with the response options set in a 5-point Likert scale from strongly disagree (1) to strongly agree (5). As explained earlier, this format was chosen in light of factors including the scale of the research, respondents' abilities and the researcher's limited time.

Table 3.1 illustrates the relationship between questions and motivation theories. The second set of questions was designed on the basis of the framework of FA (Figure 2.3).

Table 3.1 Relationship between motivation theories and motivation questions

Aspect	Theory	Questions
Language	Value	1. English is important to me 2. English is interesting to me 3. English is useful to me

		4. I am willing to offer something (e.g. time) to learn English
	Goal setting	5. Learning aims are easy to me
		6. Learning aims are specific to me
Learner	Goal orientation	7. I learn English because I want to be better than others
		8. I learn English because I want to master it
	Attribution	9. I have the ability to learn English well
		10. I work hard in English learning
	Self-efficacy	11. I have high faith in my English
		12. I am strongly confident in my English
	Achievement	13. I need to succeed in English learning
		14. I want to succeed in English learning
		15. I like the feeling of being successful in English learning
		16. I am afraid of failure in English learning
Learning environment	Extrinsic factors	17. I learn English because of my English teacher
		18. I learn English because of my parents
		19. I learn English because of friends
		20. I learn English because of school requests
		21. I learn English because I like western culture

A pilot study (Appendix 3) was conducted before the main study, to test the quality of the questionnaire. In the pilot study, 49 students from one class answered the questionnaire. Table 3.2 presents the boy-girl ratio in the pilot study.

Table 3.2 Gender distribution in Pilot study

		Frequency	Percent	Valid Percent
Valid	boy	23	46.9	46.9
	girl	26	53.1	53.1
	Total	49	100.0	100.0

The reliability of a questionnaire indicates the extent to which it is free from random error. It represents the degree to which an instrument will generate similar results in similar situations, for the same individuals at different times. Evidence of good reliability will reinforce the trustworthiness of the findings and ensure that the wider scientific community will accept the hypothesis and conclusions. High test-retest correlations indicate a more reliable instrument (Pallant, 2013). Reliability also includes internal consistency of the instrument, which is usually measured using the statistic Cronbach's coefficient alpha. Internal consistency or internal reliability refers to whether respondents' scores on any one indicator tend to be related to their scores on the other indicators (Bryman, 2016). The values range from 0 to 1, with higher values showing greater reliability. Nunnally (1978) recommends a minimum level of .70. George and Mallery (2003) also agreed that Alpha coefficients above .70 would be considered acceptable. The reliability value of the questionnaire on motivation in the pilot study, as measured by Cronbach's Alpha, is .769. This level is acceptable but there is scope for improvement.

As can be seen from Table 3.3, the reliability of the questionnaire on motivation will be increased to .804 if question number 20 is deleted. Therefore, this item was deleted from the questionnaire used in the main study.

Table 3.3 Reliability in pilot study

	Cronbach's Alpha if Item Deleted

Motivation Question 1	.743
Motivation Question 2	.740
Motivation Question 3	.753
Motivation Question 4	.751
Motivation Question 5	.763
Motivation Question 6	.749
Motivation Question 7	.762
Motivation Question 8	.758
Motivation Question 9	.753
Motivation Question 10	.740
Motivation Question 11	.740
Motivation Question 12	.740
Motivation Question 13	.756
Motivation Question 14	.760
Motivation Question 15	.738
Motivation Question 16	.774
Motivation Question 17	.771
Motivation Question 18	.772
Motivation Question 19	.751
Motivation Question 20	.804
Motivation Question 21	.731

Following five short informal conversational interviews with learners, some changes were made to the questionnaire on FA.

First of all, the arrangement of questions in this part was not ideal. To make them clearer, the questions were re-arranged into two sections: 'The use of FA in English learning' and 'Your Perception of FA'. 'The use of FA' referred to the frequency of different FA strategies, from 'never' to 'always'. Considering that changes to achievement levels were to be measured with the English quizzes, the questions in 'Your Perception of FA' encompassed only motivation. The sentence format of 'a certain FA strategy can motivate me in English learning' was used in all the questions. The responses ranged from 'strongly disagree' to 'strongly agree'.

Secondly, learners were uncomfortable with the question "My classmates always behave negatively, such as laugh at me, if I fail to answer the question". This question was selected to test whether negative peer assessment would happen in the class. Participants were uncomfortable with the phrase "laugh at" and reported that it was too extreme as an example of a negative peer-assessment. Following this advice from the learners, the question was changed to "when I answer a wrong question, my classmates would give negative feedback to me".

Lastly, there was an issue with the visual layout of the questionnaire. In the pilot study, there were no horizontal borders between each item. When students circled the answers, it was not easy for them to match the answer with the corresponding item. Therefore, lines were added to separate each item with its answers.

In summary, the questionnaire on motivation and FA was improved on the basis of reliability tests and follow-up interviews. After the amendments, all the questions were arranged on a

two-sided sheet of A4 paper with clear layout and consistent print style (Bryman, 2016; Lewin C. , 2005). The final questionnaire was ready to put into practice in the main study.

3.4.2 English Quiz

The students' achievement in the current study was measured with the English quiz. Because the fieldwork period comprised only 10 weeks, the priority was to capture changes in achievement in a relatively short time. Therefore, the quizzes were conducted weekly with the same format each week.

The quiz content covered most of the knowledge that students should have learned in each unit. A unit was a theme-focused chapter which included a passage or a conversation; a vocabulary section; a speech pattern section; a drill section with new words and new speech patterns; an extension section by doing games, surveys or role-plays; a pronunciation section; an 'inspiration' section with stories, rhymes or songs; and a self-assessment section called 'ticking time'. Each unit was designed to be taught within a week. Therefore, the items in each quiz were chosen from the unit's content.

There were four parts in each quiz, which reflected the four skills in language learning: listening, writing, reading and speaking. An example of a quiz is attached in Appendix 4. All the items in listening and writing were extracted from the teacher's book - a reference book that teacher use along with the textbook. In the reading part, the students were asked to read the passage or conversation in each unit; and in the speaking part, the topics were the themes of each unit. These selections helped ensure the validity of the quizzes because every item was closely related to the knowledge that students learned in each week.

Marking for each part was different. There were certain answers in listening and writing, but not in reading or speaking. Therefore, listening and writing were marked by the researcher according to the answers provided in the teacher's book. Reading and speaking were marked both by the students themselves and their desk mates. The researcher used the average values as the scores in these two parts. Marking criteria were provided for students on the quiz sheet. All the scores from each part were added up and converted into a scale of 10.

To test whether achievement was improving or deteriorating, a baseline score was needed. The average score from three examinations prior to the fieldwork was used as the baseline achievement score in English learning. The scores were converted into a scale of 10, to match up with the quizzes that followed.

With this quiz design and calculation method, the quizzes can record achievement in English learning during the 10-week fieldwork period, looking at different aspects of language skills.

3.4.3 Semi-structured Interview

The interview guide was designed on the basis of FA strategies that were summarised in the literature review. The interviews started with a welcome and opening which explained confidential matters and asked for permission to record the interview. This was followed by

a brief introduction to the topics that would be discussed in the interview. The interviews ended with a pilot question requesting additional information about FA in classroom, and finally an expression of gratitude for participation. The interviews were conducted in Mandarin, as it was the language that allowed interviewees express themselves freely.

Pilot interviews with two groups were conducted before the main study, in order to test how well the interview flowed and to gain some experience in that particular environment. A group interview format could help make interviewees feel safe and comfortable; it also saves time. Each group had three students, from higher-achieving, medium-achieving and lower-achieving groups respectively. There were three girls and three boys in total. None of the interviewees were informed which achievement group they were in. During the pilot interviews, three problematic issues emerged, and these were addressed and improved for the main study.

First of all, the students were nervous. It was probably because such an 'interview' was something new to them, or the 'interviewer' was new to them. Therefore, in the main study, interviews were not conducted until the fourth week (Table 3.7). In the first three weeks, the main work was to get acquainted with the students.

Secondly, students from higher-achieving and medium-achieving groups participated more actively in the interviews than the students from the lower-achieving group. Even though participants did not know they were selected according to their achievement levels, the imbalanced discourse among students was still obvious. Therefore, for the main study, the one-to-one interview format was employed instead of the group interview. This enabled all 12 student interviewees and two teacher interviewees to have equivalent opportunities to speak.

Lastly, it emerged that parents' feedback had considerable influence on students' learning, because participants from two groups mentioned the encouragement from their parents in response to the question "Any other FA strategies that had impact on your motivation and achievement were missed out in the interview". Unfortunately, the current study only explores classroom-based FA. Influence from parents' feedback could be a good topic in future research work, especially in Confucianism culture.

The interviews took place after a period of time in the field after which I had gained the teachers' and students' trust. The interview for teachers (Appendix 5) began with introductory questions on FA, such as: *What does Formative Assessment mean to you*; while the interview for students (Appendix 6 and its translation Appendix 6.1) started with direct questions on each FA strategy, such as: *Does this strategy have impact on your motivation in language learning and why*. The questions covered five strategies: sharing learning aims, questioning, feedback, peer-assessment and self-assessment. In each topic there were two sub-questions: the impact of the strategy on motivation and achievement. All the questions examined students' and teachers' understanding of FA, the use of FA in practice, and their perspectives on the impact of FA on motivation and achievement.

There are several kinds of interview questions, summarised by Bryman (2016): introducing questions, follow-up questions, probing questions, specifying questions, direct questions,

indirect questions, structuring questions, silence and interpreting questions. Each type of question was used in the appropriate area of the interviews. Direct questions were used more at the beginning with students due to the short interview duration, then follow-up and probing questions were applied to explore students' thoughts in more depth. More introducing and indirect questions were used with teachers because they have a holistic perspective of FA and they have more time to answer the questions in the interview. The interview guide listed the questions that could be used during the interview; however, if the interviewees could provide enough information to answer the research questions, not all the questions needed to be asked.

Finally, a note about recording methods. The researcher chose digital audio-recording over note-taking, to ensure fidelity of the information extracted and because taking notes *in situ* would draw the interviewees' attention to the central issues of concern for the researcher.

3.5 Population, participants and sampling technique

This research studied the impact of FA on primary school students' motivation and achievement. The population in this research was a group of Chinese primary school students in Jiangsu Province who were learning English. In most of the public primary schools in China, students start to learn English from third grade when they are about nine years old. This research studied the students in fourth grade, when they have been learning English for one year, because at that stage, foreign language class would be a normal school subject to them, just like maths and Chinese.

Having a mixed methods design, this study adopted mixed methods sampling. With it being a concurrent study, where quantitative and qualitative methods were conducted separately but integration of data would occur at the data analysis stage only, multilevel design was applied. Participant samplings were at four levels:

1) The schools – convenience sampling.
2) The teachers – random sampling.
3) The classes who fill in the questionnaires and quizzes – once the teachers were selected, the students were automatically selected.
4) The students who took part in the interviews – stratified purposeful sampling.

Table 3.4 summarises the number of participants in the pilot study and the main study. Students in the pilot study all signed the consent agreement, while in the main study, 196 out of 235 students in the sampled classes agreed to take part in the study and signed their consent forms.

Table 3.4 Number of participants

Research methods	Instruments	Pilot study		Main study	
		Teacher	Student	Teacher	Student
Quantitative methods	Quiz/Questionnaire	/	49	/	196
Qualitative methods	Interview	1	6	2	12

Convenience sampling is the easiest technique to gain access to subjects. It saves researchers' time, effort and money. However, it has the lowest level of credibility of all the qualitative sampling designs, as it is neither purposeful nor strategic (Gray, 2014). Considering the scale of this study, the time of the fieldwork, and the researcher's practical constraints, convenience sampling at this level was the best, pragmatic choice. Therefore, two primary schools in Jiangsu Province were chosen because the head teacher had a personal relationship with the researcher. These two schools belong to one education group. Teachers from these two schools gather together every week to prepare the lessons, hence all the material and preparation for teaching were the same.

The research required a great deal of support and cooperation from the teachers. Therefore, teachers were the next sampling level to be considered, rather than the class. Within the group of fourth grade English teachers, one was willing to participate in the pilot study and two were willing to participate in main study, so these teachers became the subjects in the study and all their students became the eligible participants.

Students participating in the interviews were the only group of participants that researcher had the chance to select differently. The sampling technique applied here was stratified purposeful sampling. This technique is able to achieve a greater degree of representativeness and reduce the degree of sampling error (Gray, 2014). Each stratum is constituted by a group of participants with the same characteristics - in this case, the achievement level. This approach enabled the comparison of the major variations across strata (Gray, 2014; Cohen, Manion, & Morrison, 2011).

In the pilot study, two students were randomly selected from each achievement level, i.e. six students in total. They were randomly assigned into two groups. Two group interviews were conducted in the pilot study. Similarly, two students were randomly selected from each achievement level in each FA environment, i.e. 12 students in total. Twelve individual semi-structured interviews were conducted in the main study. Table 3.5 summarises the sample of student interviewees in the main study.

Table 3.5 Student interviewees sampling

FA environment	FA – rich environment						FA – lacking environment					
Gender	B	G	B	G	B	G	B	G	B	G	B	G
Achievement level	H	H	M	M	L	L	H	H	M	M	L	L

H: Higher achievement level
M: Medium achievement level
L: Lower achievement level

B: Boy
G: Girl

As mentioned earlier, Chinese primary school students learning English were the community that this research targeted. This section has discussed why students in fourth grade were chosen, and why and how different sampling techniques were applied at different levels. The following section will elaborate the process of collecting data.

3.6 Procedure and data processing

To explain the research procedure clearly, the table below shows the timeline including the pilot study and the main study.

Table 3.6 Timetable of the study

Phase of study	Weeks	Date: Week Commencing (Monday)	Description of work	Who with		How long is the work
Pilot study	1	02/03/2015	Semi-structured interview	Teacher	1 teacher	30 minutes with the teacher
			Focus Group	Student	2 group interviews with 6 students: HG, MB, LG; HB, MG, LB.	20 minutes for each group interview
			Pilot questionnaire		49 students from the class	10 minutes
			Quiz		49 students from the class	20 minutes
		09/03/2015	Analyse Pilot study data and adjust research instruments			
Main study	1	16/03/2015	Questionnaire	Student	235 students from 4 classes	10 minutes
	2	23/03/2015	Quiz		235 students from 4 classes	20 minutes
	3	30/03/2015	Quiz		235 students from 4 classes	20 minutes
	4	06/04/2015	Quiz		235 students from 4 classes	20 minutes
			Semi-structured interview		HBR HBL	20-25 minutes
	5	13/04/2015	Quiz		235 students from 4 classes	20 minutes
			Semi-structured interview		HGR HGL	20-25 minutes
	6	20/04/2015	Quiz		235 students from 4 classes	20 minutes
			Semi-structured interview		MBR MBL	20-25 minutes
	7	27/04/2015	Quiz		235 students from 4 classes	20 minutes
			Semi-structured interview		MGR MGL	20-25 minutes
	8	04/05/2015	Quiz		235 students from 4 classes	20 minutes
			Semi-structured interview		LBR LBL	20-25 minutes
	9	11/05/2015	Quiz		235 students from 4 classes	20 minutes
			Semi-structured interview		LGR LGL	20-25 minutes
	10	18/05/2015	Quiz		235 students from 4 classes	20 minutes
			Questionnaire		235 students from 4 classes	10 minutes

			Semi-structured interview	Teacher	2 teachers	45 minutes interview with each teacher

3.6.1 Pilot Study

As stated earlier, one teacher and 49 students participated in the pilot study. The pilot study lasted for one week. Quantitative data were collected in the form of a questionnaire and an English quiz; qualitative data were collected via a semi-structured interview with the teacher and a group interview with six students.

In the pilot study, at the beginning of the week, consent forms were distributed to the students. As they were not adults, the forms were taken back home to be signed by their guardians. Every child was willing to participate in the study. Once all required consent forms were obtained, the data collection could begin.

In terms of quantitative data, questionnaires were distributed to the students the day after gaining consent. The students were asked to complete the questionnaire in the classroom, then five of them were randomly selected to comment on the questionnaire. At the end of the week, the students were given an English quiz which covered the knowledge being taught that week. The quiz papers were collected and marked by the researcher.

In terms of qualitative data, two group interviews were conducted during the week. On the basis of the students' previous performance in tests, they were grouped into three achievement levels. The group interviews were conducted with three students from each achievement group. Besides this, an individual interview was conducted with the teacher. After each interview, interviewees were asked to comment on the questions that were asked, whether the questions were appropriate, and whether any information might be missed in the interview.

One of the aims of the pilot study was to adjust the research instruments. With the feedback from the participants and the reliability test of the questionnaire, the instruments were modified (3.4) and were ready to use in the main study.

3.6.2 Main Study

Two teachers and 235 students participated in the main study. With one teacher having more of a grounding in FA than the other, the difference between two FA environments was potentially significant and a comparison between the performance of the two cohorts could be made to examine this. The main study lasted for ten weeks. Quantitative data were collected in the form of a questionnaire and an English quiz; qualitative data were collected in the form of semi-structured interviews with the teacher and 12 students.

In the main study, the consent form was given to the students at first and they took them home to their guardians. There were 235 students in the four classes led by the two teachers (two classes each). A total of 196 students and their guardians agreed to participate in the study.

The research procedure for the main study that compared the two cohorts - FA-rich environment (FAR) and FA-lacking environment (FAL) - is outlined in Table 3.7

Table 3.7 Research model and procedure of the current study

Week	1	2	3	4	5	6	7	8	9	10
FAR	Questionnaire Pre-test	Quiz	Quiz	Quiz Interview	Quiz Interview	Quiz Interview	Quiz Interview	Quiz Interview	Quiz Interview	Quiz Interview with teacher Questionnaire Post-test
FAL	Questionnaire Pre-test	Quiz	Quiz	Quiz Interview	Quiz Interview	Quiz Interview	Quiz Interview	Quiz Interview	Quiz Interview	Quiz Interview with teacher Questionnaire Post-test

In terms of quantitative data, the modified questionnaire was administered in the first week and again in the tenth week. They were referred to as the pre-test and the post-test thereafter. The researcher briefed the participants about her background, the aim of the study, the content of the questionnaire and the time it may take. The students were asked to read the instructions and had the opportunity to ask questions before completing the questionnaire in class. A few students asked in the post-test why they needed to do the same questionnaire again. The researcher explained that the post-test was used to compare their performance in the pre-test, as the questionnaire was coded with student's registration number. There were nearly ten weeks between the pre-test and the post-test. It was unlikely for students to remember their answers to the questionnaire and fill in the same answers in the post-test. Therefore, the changes in their motivation, the use of FA and students' attitude towards FA can be shown in the comparison between the pre-test and the post-test.

Another set of quantitative data was collected weekly in the form of the English quiz. Quizzes were distributed at the end of each week, when the students had learned the whole unit for the week. As mentioned before, each quiz was produced by collecting the items in the teacher's book in every unit. The quiz could therefore test how well the students had mastered the content in each unit. There was a 20-minute reading session every Friday morning and the teachers kindly made this time available for the quiz. As noted above, the quiz was designed to cover four skills areas – listening, writing, reading, speaking. It normally took students 10 minutes to finish the listening and writing parts. The next 10 minutes were used to complete the reading and speaking parts. Sometimes, not every student had a desk mate for peer marking of the reading and speaking parts. If students were sitting on their own, they were asked to join with the two students in front of them and mark each other within the group of three.

In terms of qualitative data, 14 individual semi-structured interviews were conducted from the fourth week with the 12 students specified in Table 3.6 and the two teachers. Because of the limited free time that students had in school, most of the interviews could only be done during the 20-minute breaks in the classroom. If there was an eye-protection exercise, another five minutes were used. There were two problems with this arrangement: time and interview environment. Time was limited, and some interviews were interrupted by the class bell. The student interviewees were called to the interview spot right after the teacher dismissed the lesson and stayed there until all the interview questions were asked.

Subsequently, students were facing extra pressure because of the sudden loss of their break time and the unexpected interview. Fortunately, most of the interviews were finished within 20 or 25 minutes. To help ensure that students were calm, the interviewer also asked their consent to continue the interview and emphasised that they had the option to drop out of the interview at any time, and they could speak freely in the interview. The interview environment was sometimes challenging because interviewees were always surrounded by other students when being interviewed. Although the location for interviews often involved a move from the student's seat to a corner in the classroom, the stairs, or even the yard, the environment in the school was naturally noisy and crowded due to the break time. However, the break time was the only time that the interview can be taken place because it was unethical and inappropriate to ask students to skip other lessons. There was no practical solution to these problems in the current study.

Apart from these logistical problems, all the interviews went smoothly. First, students were asked for permission to be audio-recorded. Because the researcher had been involved in their daily English teaching and learning for three weeks, interviewees were comfortable to share their thoughts with the researcher. Questions were asked in accordance with the interview guide, and all the answers were recorded. Two incidents of note occurred. One of the interviewees from the lower achievement group asked why he was selected to be interviewed. To protect the interviewee's self-esteem, the researcher did not tell him about the achievement group, and only said he was chosen randomly. Another girl, also from the lower achievement group, did not wish to be recorded with a camera. Therefore, in that interview, no image was recorded. Both students were from the lower achievement group and it is possible that lower-achieving students were more sensitive than others to certain issues. These incidents had the effect of reminding the researcher to be sensitive when working with these students.

The two interviews with teachers took place in their offices. There were no time limits for their interviews and the offices were quiet and private. They were free to express their opinions during the interview. Both teachers talked for about 45 minutes. At the beginning of the interviews, routine questions about consent were asked. Then the interviews proceeded in accordance with the interview guide (Appendix 5).

This section has explained where, when and how the data was collected. The following section will explain how each data set was analysed, e.g. questionnaire, quiz and interview. Coding and analysis of data will be elaborated in Chapter 4 and Chapter 5.

3.7 Data analysis

The analysis consisted of three steps, preparing, processing, and analysing. There were three data resources in this study: questionnaire, quiz, and interview. The interpretation of data differs due to the nature of the data. In the following sections, each step of analysis for different data will be explained.

3.7.1 Preparing data

In the preparation, the researcher can familiarise herself with the data and ensure that the data is complete and accurately recorded. Meanwhile, it would be a good opportunity to organise the data in order to get it ready for the further steps (Matthews & Ross, 2010).

The quantitative data was collected from questionnaires and quizzes. All the data was entered into SPSS including numerical coded personal information. For example, the boys were coded as 1 and girls were coded as 2; FA rich environment was coded as 1 and FA lacking environment was coded as 2. The category codes were used to compare the data between different groups.

The quiz results were numbers scaled from one to ten. However, the baseline, which was calculated from three previous exams, was in the scale of 100. Therefore, after divided by ten, the baseline scores were entered into SPSS too.

Some data was missing due to the inevitable practical reasons, such as students missing the quiz, or not answering some questions in the questionnaire. Hence, the missing values were defined as 0 (zero) to continue the data analysis. Zero has no other meaning in the coding system in this study, therefore, there was no problem in coding data as 0 (zero) technically (Robson & McCartan, 2016).

The qualitative data was collected from interviews. The first step to prepare qualitative data was transcribing and entering to NVivo. The raw data was visual data which consisted of video and audio resources. There are several good reasons why the social researcher might use visual data in their studies, one of them is that visual data may reveal some sociological insight that is not accessible by any other means (Banks, 2007), also, the data presentation via images may be more direct and powerful. This is why the data was collected via video recorder in the first place. However, due to the ethical protection, the images in data presentation was not considered because all the student participants were under 18-year-old. The interviews were conducted in Mandarin, therefore, the initial transcript was in Chinese. 14 individual transcripts were inputted into NVivo, and were translated into English in later phases. More details about translation will be discussed in 3.8.1.

3.7.2 Processing data

Before any data test, the quality of the questionnaire was examined via Reliability Analysis and Correlation Analysis to ensure that the questionnaire was reliable and valid (Bryman & Cramer, 2011).

The questionnaire data consisted of three parts: the scale of their motivation in language learning, the use of FA and their attitude towards the FA strategies. The variables concerning motivation were tested in Factor Analysis to decrease the number of variables (Bryman & Cramer, 2011), and determine the main motivation factors in learning language. New variables about motivation factors were generated, and the mean value of each factor was calculated from grouped motivation questions according to the Factor Analysis results. All the data about the use of FA, students' attitude towards the FA strategies and quizzes results would be compared to the motivation factors rather than each motivation question.

The students were grouped into three groups based on their achieving levels. The cut-points were defined when the proportion of each group was stable during the fieldwork. Three groups of students at different achieving levels were compared in terms of their motivation in learning language, FA use, and attitude to FA. Furthermore, the triangular correlations between motivation, FA and achievement were compared between different achieving groups. At the same time, the average achievement was calculated using the mean score of ten achievement results. It was seen as the general achievement for each participant.

In the data preparation, 14 interviews were transcribed individually. Meanwhile, in hardcopy, 12 student interviewees' answers to each question were listed vertically and all the answers were arranged in the order of interview guide (Appendix 6). For example, 12 answers to interview question 'do you like the teacher sharing learning aims with you and why' were excerpted and listed in a column. Horizontally, it was all the answers to every question from one interviewee. This technique can merit further differentiation (Dey, 2005) because all the information was easy to be compared between interviewees as it was on one piece of giant paper.

In the first two steps, all the data were read again and again, the holistic image of data was pictured. The following step was to pull all the fragmented information together and make connections between different ideas.

3.7.3 Analysing data

The descriptive analysis was conducted firstly, including the ten achievement from quizzes; the pre-test and the post-test data of motivation in learning language, the use of FA strategies, and the attitude to FA. The students were divided into different groups in the light of gender, FA environment and achievement level. Each set of data was explored via SPSS, compared between groups and presented in the form of diagram.

To explore the impact of FA on motivation and achievement, the relationship between them needed to be defined. Therefore, the relationships between the pre-test/post-test FA, the pre-test/post-test motivation and the mean score of achievement were analysed between each two via Bivariate Correlations (Bryman & Cramer, 2011). This explained the strength and the direction of the relationship between each pair of variables. In all the correlation tests, the relationships were compared between groups, which included gender, FA environment and achievement level. All the significant correlations were highlighted and explained in Chapter 4.

The interview transcripts were coded electrically in NVivo, using coding strategies such as repetitions and regularities; compare and contrasts; compare with hypothetical or extreme examples (Bazeley & Jackson, 2013). They were also coded manually on aforementioned giant paper. The codes from two recourses were compared and emerged in later phase. There are two approaches to coding, from general categories to details, or the opposite (Bazeley & Jackson, 2013). In this study, the transcripts were coded in detail first. All the codes were rearranged, renamed and recategorised into a hierarchy structure (Appendix 12.2) after the two recourses were compared and emerged. Appendix 12.3 depicted the hierarchy diagram of codes that compared by number of coding references.

In this section, how each set of data was analysed was explained, and the next section will discuss the inevitable ethical issues and cultural considerations.

3.8 Ethical and cultural considerations

Ethics is not a part of the research process that can be dismissed. Social research is about human beings, not only because it involves human beings directly, but also because the researchers, as human beings, will have an impact on the participants, the research, and even themselves (Matthews & Ross, 2010). Ethical considerations are vital when addressing conflicts between the law, the rights of individuals and research designs, especially with people who were not familiar with written consent. With reference again to Clough & Nutbrown's cookery metaphor, ethical considerations ensure that the recipe and the ingredients used are safe and harmless to the people who help to make the dish. In this section, I discuss cultural considerations and ethical issues, as well as improvements or solutions to dilemmas.

3.8.1 Doing research in primary schools in China

Doing fieldwork in China

Being a Chinese researcher was definitely an advantage for doing research in China. However, it would be unfair to take some issues for granted as an 'insider' but not declare them. People from other cultures, for example western societies, may not have the same understanding of Chinese ways of responding or expression. Discussion of culture and language issues helps provide audiences with a better insight into the way potential losses in meaning have been avoided to the maximum extent (Squires, 2009).

First of all, the current study was carried out in a cross-language environment. In the research paperwork, language was translated twice in this study: in the questionnaire, survey, and interview questions preparation (English to Chinese) and in data for presentation (Chinese to English). The theoretical frameworks were summarised from Anglophone-dominated publications. The motivation framework was translated into Chinese and used for the items in the questionnaire. In data collection, Mandarin, the language that participants used in their daily life, was used all the time. The only exception was the information sheet for teachers. The teacher-participants were English language teachers, who were able to read and fully understand documents in English. The contents of the information sheet were explained orally in Mandarin too. Using English in the information sheets therefore did not affect the data. On the contrary, it could be argued that using English in the documents benefited the procedure by enhancing teachers' trust in the researcher due to these formalities. In analysis, qualitative data were analysed in a Chinese environment and translated into English for presenting and reporting. There was little impact of language on the quantitative data collected from questionnaires since these data were in numeral format. However, it is acknowledged that the process of translating items in the questionnaire may have produced some subtle, nuanced differences between the English version and Chinese version.

As a bilingual researcher who has been studying and teaching in the UK for five years, I acted as my own translator with the help of computer-assisted translation platforms. Translation is important in cross-language research and it is not a simple technical procedure but a complex social and cognitive process which can affect outcomes (Shklarov, 2007). Therefore, the first stage was to translate language with the aid of mature, intelligent computer-aided translation platforms (Google, Youdao, iCIBA). The second stage was to adjust the translations manually. The adjustments included correcting the grammatical mistakes; completing the sentences by adding subtext, for instance, subject; re-interpreting on the basis of culture and language differences, *etc.*

There was an undeniable risk that the meaning of data might be partially missing or distorted in the translation. There were several nuances in the translation of the questionnaire from Chinese to English. For example, in Appendix 2 and its translation Appendix 2.1. the statement about negative peer assessment was '*The negative feedback from my classmates can motivate me'*, whereas the literal translation from Chinese was '*The classmates pointing out my shortcomings makes me want to learn English more*'. Another example was in the interview transcript that one interviewee said that *'[I] still need something like a blow/stroke*'. The '*blow*' or '*stroke*' was defined as negative self-assessment, because the interviewee was talking about self-assessment. Since the interviews and questionnaire were conducted in Chinese, the English translation only appeared in data presenting. However, the nuances of the information that implied by the participants may not be interpreted in the translation. There are several methods to minimise the gap between different languages (Nes, Abma, Jonsson, & Deeg, 2010), for example: focus on the thinking and reflection during the analyses; stay in the original language as long and as much as possible; and have discussions with a peer who does not speak the source language. All of these methods were used to avoid or minimise losses of meaning.

There are some advantages of being a translator for one's own work. As the person who designed and planned the research, the translator understood the research the best and would not misinterpret the language of method and procedure. As the fieldworker and data gatherer, the translator could spot and analyse the nuances in interviews and preserve them in translation. As a person who understood both cultures, the translator could pay attention to the cultural differences and preserve diversity.

On the subject of cultural diversity, the fundamental philosophies in China and in the UK are different. Confucian philosophy, which shaped the Chinese cultural values from the Han dynasty (206 BC-220 AD) onward (Jacobs, Gao, & Herbig, 1995), is central to Chinese interpersonal behaviour.

Confucianism's tenets were harmony, hierarchy, and developing one's moral potential and kinship affiliation, which together provide the foundation for human relationships (Shenkar & Ronen, 1987; Carless & Lam, 2014). Humans exist in relationship to others (Cui, 2015). Chinese people call the relationships *Guanxi*, a Chinese term which refers to interpersonal connections (Chen & Chen, 2004). Therefore, the *Guanxi* between students, teachers and researchers should be taken into consideration when analysing participants' responses. This is because the research outcomes may be affected by Chinese philosophical beliefs, such as

trying to keep the harmony in the relationship, venerating and deferring to elders in this hierarchical society (Matthews & Ross, 2010).

At the same time, the researcher benefited from *Guanxi* of kinship. During the fieldwork, the first task was recruiting research participants. Thanks to interpersonal networks, I was introduced to the head teacher of my old primary school, who was my own teacher when I was young, as well as a friend of my parents. She provided her full support after I explained my research design and potential ethical problems. The regulations, such as application for access to schools and children, were 'flexible' owning to *Guanxi*. For the same reason, I was introduced to two teachers who were willing to take part in the research. To the teachers, I would be either an 'insider', a Chinese who could make people relax in ways they would not have if I was a non-Chinese, or an 'outsider', a person sent by the head teacher who was their immediate boss. To minimise the psychological distance, I told the teachers to treat me as an intern teacher, who would be happy to assist them with daily work, but who needed to collect data from them and the students.

Lastly, the impact of the One-Child Policy was another unique phenomenon in China. Unlike the UK, most Chinese people born after 1979 were the only child in their families (Zhu, 2003). They were called 'little emperors' because they were regarded as significantly less trusting, less trustworthy, and less conscientious (Cameron, Erkal, Gangadharan, & Meng, 2013) than previous generations. The reasons behind this phenomenon were the structure of the family in China and the dotage from older generations. In most of the one-child families, there would be four grandparents, two parents and one child. This 4-2-1 inverted pyramid structure tends to increase the weight of expectation on each child (Dello-Iacovo, 2009). With no other children in the families, the needs of the only child were usually fulfilled immediately without considering others' interests, no matter in the family or in the community. But as time passed, the only-child generation gradually turned into a selfish generation, according to observers (Hvistendahl, 2013). Some of the participants' perceptions of FA or aspects of their motivation can be traced back to this unique atmosphere for this generation, created by the conflicts between the traditional collectivism in Confucianism's culture and the new individualism.

Doing research with children

Children bear two identities: human being and child. Children are not mere subordinates to adults. As human beings, they deserve basic respect. According to the "Research ethics: General Principles and Statements" guidance from The University of Sheffield, the paramount principle is "respect for the participants' dignity, rights, safety and well-being". Principles of biomedical ethics, first published in 1979, also illuminate the four fundamental tenets: autonomy, non-maleficence, beneficence, justice (Greig, Taylor, & MacKay, 2012). The principles include participants' free choice, no physical or non-physical harm, awareness of risks and benefits, and fair and equal treatment. The primary rule in this research was never to violate anyone's human rights or human dignity.

However, children are not simply immature adults. They have special natures which are different from adults and which require special sensitivity from the social researcher (Greig, Taylor, & MacKay, 2012). They are special and highly susceptible to external influences and

factors. Therefore, it was necessary to take potential external factors into consideration as much as possible. The trustworthiness of data generated from children should be weighed against truth value, applicability, consistency, and neutrality (Shumba & Chisaka, 2013).

Personally, as a researcher who is also an educator with a degree in primary education, I brought ample relevant skills and experience of working with children. I had received training in educational psychology, child psychology, classroom management, *etc.* I had gained a lot of practical experience of working with children during intermittent blocks of teaching and learning over the past nine years. I also obtained the Disclosure and Barring Service (DBS) check in the UK.

In summary, it is clearly vital that the research takes account of the characteristics of the Chinese participants: as children, as adults, as human beings, and as members of a Confucian culture.

3.8.2 Ethical issues and countermeasures

Ethics was not an 'add-on' principle (Greig, Taylor, & MacKay, 2012). From research design and implementation, to analysis and presentation, researchers need to reflect on the ethical issues raised by their research and need to justify the practices and procedures. When planning the research, it is crucial to prepare for all possible incidents and the corresponding risk management.

Research design

The very first research design conceptualised by the researcher involved asking the teachers to use more/less FA in one of their classes while keep the FA-lacking/FA-rich situation in another class. However, in such a scenario the children would not be treated fairly or ethically. The children in the FA-lacking environment were denied the right to receive more FA than they could have received; and the children in the FA-rich environment received assessment which was less than before. This plan was therefore rejected by the researcher. With the current research design, there were no significant changes in use of FA by each teacher, possibly leading to a less notable change in students' motivation and achievement. However, this was the best compromise proposal in light of ethical considerations.

Diversity of participants

Diversity in subjects was another factor that was considered. The target participants for this study were primary school students who were learning English in China. Fortunately, the numbers of boys and girls in the sample were roughly equal. Every student in the class participated in the study, filled in the questionnaire and survey, but not all the data was recorded due to the difficulty in gaining the consent. The participating students fitted a normal distribution model of achievement. Those who were invited to attend interviews were selected from three groups: higher-achieving, intermediate-achieving and lower-achieving. They were informed that they were chosen because the researcher wanted their help with the study. The more specific reason was not divulged in order to protect the self-esteem of lower-achieving students and avoid influencing responses to interview questions.

Informed consent, anonymity and confidentiality

Ethical consideration was planned and given according to the needs of the current study. The consent was obtained twice during the fieldwork.

Firstly, before questionnaires were distributed, the volunteers were informed of their obligations and commitments and those of the researcher. Clear statements were printed out on the information sheet and consent form; verbal explanations were also given about why the research was conducted, how the answers and volunteers' identities would be protected, and what risks there were in the procedure. The information sheet and consent forms were drafted and redrafted and approved by the University of Sheffield University Research Ethics Committee before going to the field (Appendix 7). The examples of the information sheet for students (Appendix 8 and its translation Appendix 8.1), and for teachers (Appendix 9), and the consent form for students (Appendix 10 and its translation Appendix 10.1), and for teachers (Appendix 11) are attached in the Appendices.

Secondly, before recording the interviews, consent issues were stressed again. Interviewees' answers would not be divulged and their real names were coded with pseudonyms in the analysis and presentation of data. Because all the student participants were under 16 years old, written consent was specifically needed. Therefore, children were asked to take the information sheet and consent form home to be signed by their guardians.

Gaining consent from people to be included in the research was not just about signing the consent form. Because the people in the fieldwork environment were not familiar with the concept of 'doing research', it was more difficult than expected to gain permission from the participants as well as the gatekeepers. Therefore, the terminology and the processes of research were explained carefully and verbally throughout the fieldwork until it was clear that gatekeepers and participants understood their rights and obligations, particularly anonymity and confidentiality in the study. As already stated, written consent was obtained from every participant and oral agreement from interviewees. They were reassured that any trace of their identities would be removed from any piece of the interview transcript or recording.

Power imbalance

There was potential for power imbalance in the researcher-participant relationships. This may be hard to avoid especially when there is a vulnerable group in the research. Primary school students, as well as their parents, living in a second-tier city in China, may find the idea of participation in research unfamiliar, unnecessary, and uncomfortable (Habibis, 2010). 'Doing research' was something beyond their knowledge. Therefore, the first step in fieldwork was to be a 'neutral' person in their classes at school and weaken the impact of 'doing research' to help redress any power imbalance. My roles were akin to that of a big sister to the students and an intern teacher to the teachers, who sometimes would talk to them to gain information which would contribute to Chinese education. Once all the participants were showing no reticence and had got accustomed to having the researcher in class and talking to them and recording the conversations, the power relationship was

balanced to the greatest extent possible in the circumstances.

Data management and storage

All the raw data was coded without personal information which could link back to the participants. (This is why the Appendices have a copy of the electronic information sheet and consent form rather than signed ones.) All the personal information was disposed of once data analysis was complete.

3.9 Reflection and reflexivity

This study used many methods to improve the quality and validity of the results, such as combining quantitative and qualitative techniques to produce comprehensive data and findings, and taking all relevant variables into consideration in research design, *etc*. It is nevertheless true that many unexpected or inevitable issues still arose in the current study. This section will discuss the drawbacks of the research methods, as well as possible improvements in the future work.

3.9.1 Theoretical framework

Formative assessment

Five FA strategies were analysed in this study. They are sharing learning aims, questioning, feedback, peer-assessment and self-assessment. These strategies were summarised from earlier literature and were representative of the main FA strategies used in the classroom. Gathering all these strategies together presents a comprehensive picture of formative assessment. Moreover, this study took one step further to classify some of the strategies, such as specifying three categories in teachers' feedback. The author aimed to obtain a comprehensive picture but she felt that more work could have been done to differentiate aspects of each strategy. For example, sharing learning aims would be different in different types of lessons: in a new lesson, the aim could be the target vocabulary or speech pattern; in a grammar lesson, the aim could be to use the language in a possible scenario; in a review lesson, the aim could be to finish an in-class test. The aims may serve the class differently. Because of the simple classification approach, this research lacks in-depth investigation of individual strategies.

Motivation

Seven motivation theories were integrated and applied in the current study. These theories can provide a comprehensive picture of students' motivation in language learning. However, motivation is a complex concept. Twenty items about motivation in the questionnaire can capture an assessment of the participants' overall motivation. To make sure that the primary school students felt comfortable with the questionnaire, the initial plan to have more motivation questions was abandoned. The depth will be guaranteed at the cost of breadth.

Achievement quiz

Firstly, although the validity of the quiz items was ensured by the method of item selection (3.4.2), parametric examinations were still recommended if there was an opportunity because they may diminish the influence of unbalanced question items. Secondly, more question items could be tested to increase the reliability of the achievement data. In this study, to minimise the students' workload, 5 or 10 questions in each language skill were arranged on a one-sided A4 paper. To increase the number of questions while easing the burden of work on students, the quiz could be divided into four parts according to the target repertoire and each part distributed separately during the week as homework. However, this method may cause inconvenience to the English teacher. In future work, if the teachers are supportive, the reliability of the achievement data could be increased by using this alternative method. Lastly, innovative technology could be used to test reading and speaking. Due to the shortage of personnel, funding and time, the reading and speaking sections were marked by desk mates using pre-set criteria. However, the criteria do not represent a uniform standard. The students' marking may differ for these two skills. If classroom logistics allow, it would be better if all the students could record their replies and upload them to be assessed by the researcher. In this case, all the answers and replies would be assessed with unified criteria. Nevertheless, without external funding or sufficient time, peer-assessment with pre-set criteria was the most convenient and efficient plan for this study.

3.9.2 During the fieldwork

Time length

The length of the fieldwork period shrank from three months to ten weeks due to an unpredictable incident. The first ethics application was rejected by Research Ethics Committee at the University of Sheffield. One of the main reasons for this related to how the research and the researcher were understood in China, especially since the work was undertaken under the authority of a university outside of China. As mentioned in 3.7.1, thanks to the principles of *Guanxi*, recruiting volunteers was not difficult at all. Unfortunately, due to the intercultural differences which led to the ethics approval delay, the schedule of the fieldwork was disrupted. Owing to the winter holiday in China, the new schedule involved a one-week pilot study followed by a ten-week main study to ensure that a complete set of fieldwork could be completed. The revised time period produced a dilemma: it is not so short that only a 'steady picture' could be captured, but it is potentially not long enough to observe significant changes. For this study, ten weeks was the longest time that the researcher could stay in the field. Even though the duration was short, the findings showed some significant but not satisfactory results. The impact of different FA environments may emerge more clearly if the students have more time to absorb the influence of FA. A considerably longer time in the field would allow the author to collect richer data and conduct the research at an unhurried pace. It would be more conducive to putting into practice the idea of distributing the quiz as homework. Also, in practice, the students were interviewed in the breaks between lessons; it was not ideal for students not to have adequate breaks. A longer fieldwork period would create more opportunities for

interview slots. Students would have more time to express themselves freely, and there may be less influence from classmates' presence or the interview environment

Sample

There were 196 students who participated in the questionnaire, while 12 students and two teachers participated in the interviews. It was appropriate to generalise the findings at a provincial level, but due to the massive scale of Chinese schools and the discrepancies in Chinese education across different provinces, it was difficult to generalise the findings to a wider context, such as nation-wide. It is hoped that this study can present a good example of a case study of formative assessment in the context of Jiangsu Province. Education in rural areas, especially the English language education, is totally different from that in cities. The rural areas are short of teachers, teaching materials and expertise about the best teaching methods, *etc*. Considering these dramatic differences between urban and rural areas in China, more studies might be conducted in rural areas. The potential impact of formative assessment may be greater there than in cities. In other words, it could be more useful in rural areas. However, this is merely a hypothesis which needs to be explored in the future.

3.9.3 Data analysis

Apart from the issue of missing details in translation discussed in 3.7.1, there is an aspect that was neglected in the qualitative analysis: the students' physical expressions. In this study, attention was focused on the verbal information in the interviews, but little attention was given to physical expressions and non-linguistic communication, such as facial expression, body posture, gesture, tone of the voice. However, body language could have revealed more if the author had investigated the particulars in a systematic way. For example, an embarrassed smile, avoiding eye-contact, folded arms, or a tone of irony all potentially indicate a negative attitude to the interview items. In the future work, physical expression in the classroom is a promising area to study.

3.10 Summary

This chapter has discussed the philosophical stance of the author, the justification for conducting such a cohort and comparative study, the reasons for applying mixed methods, how the research instruments were designed, and how the sample was selected. To reiterate the cookery metaphor, it might be said that the chapter has explained how the recipe was devised and the ingredients chosen, in order to prepare and cook a well-crafted dish called research. Then the chapter proceeded to describe the process of conducting the research in China, as well as cultural considerations and ethics issues in this study. Lastly, the limitations of the research design and methods were discussed via reflection on the study, followed by the suggestions for future works.

The next chapter will present the quantitative and qualitative analysis and findings.

Chapter 4 Quantitative Data Analysis

4.1 Introduction

This chapter starts by presenting the quantitative data obtained from the main research investigations, by setting out the statistical analyses of questionnaire results and quiz sheet scores.

Subsequently, the quantitative data are drawn out in more detail from the questionnaire and quiz results. The questionnaire related to learners' motivation in learning language in the class; the use of FA within the current class; and their motivational attitude towards different FA strategies. The quiz sheet covered the four learning skills: listening, speaking, reading and writing, to assess the level of language knowledge mastered during every week of the investigation period. Using the statistical analysis software programme SPSS 22, the quantitative data were analysed using various appropriate statistical tests to examine the relationships between formative assessment, motivation and achievement, in answer to the research questions 1, 2, 3.

4.2 Participants

A total of 235 copies of the questionnaire were sent out to four classes. Of these, 196 students signed the consent agreement and submitted analysable questionnaires, therefore, 83.4% of the responses were accepted and analysed in this study.

The learners had a variety of ages ranging from nine to 13 years, with an average of 11.22 years. Of the 196 responses, as presented below in Table 4.1, 100 were from boys and 96 were from girls; 99 were from an FA-rich environment, led by a veteran teacher, and 97 were from an FA-lacking environment, led by a novice teacher. The veteran teacher had been teaching English for ten years. She was awarded for her excellent teaching, as well as her leadership in the city-governed teaching group. In the classroom observation, it was found that she did her job with skill and ease. This was evidenced by her skilled assessing habits, flexible teaching plan, good oral English, and so forth. The novice teacher had graduated from a university two years ago when the study was conducted. As a new teacher, she was able to deliver good knowledge of English to students. However, the teaching style was more rigid when compared with the veteran teacher. The main task in her teaching was to cover each part of the textbook in a fixed sequence. Apart from managing class discipline, not much attention was paid to react to students' performance. Therefore, in her class, the FA was lacking. The veteran teacher had a good understanding and rich practices of FA, while the novice teacher did not. However, the novice teacher was very modest and had a keen desire to make the changes to improve her teaching, while the veteran teacher was very satisfied with her teaching and did not intend to change.

Overall, for gender and FA environment, a similar percentage of learners responded (around 50.0%) from each sub-group.

Table 4.1 Gender * FA environment of participants

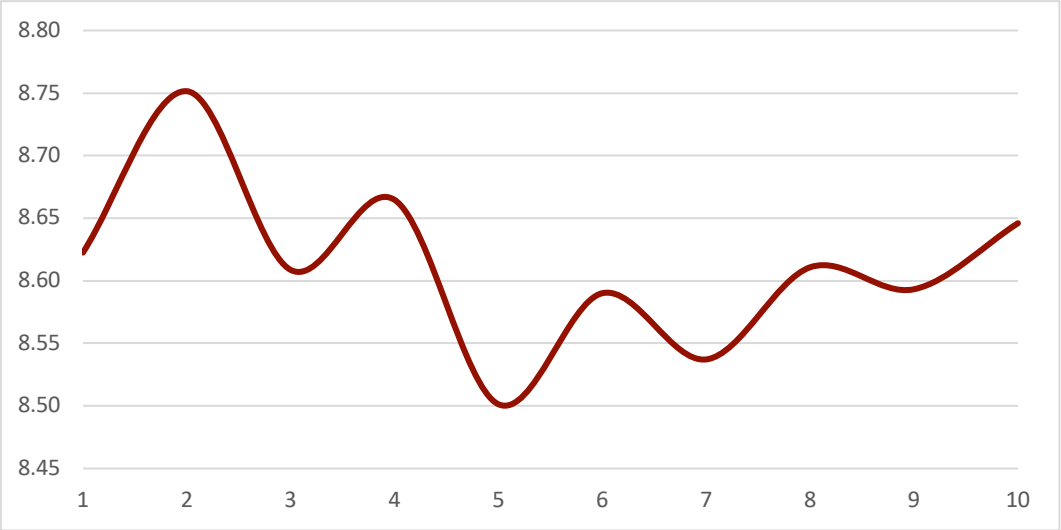
			FA environment		
			FA-rich environment	FA-lacking environment	Total
Gender	boy	Count	49	51	100
		% of Total	25.0%	26.0%	51.0%
	girl	Count	50	46	96
		% of Total	25.5%	23.5%	49.0%
Total		Count	99	97	196
		% of Total	50.5%	49.5%	100.0%

4.3 Achievement

The students' achievement in this study was measured with a baseline score and results from nine quizzes. The baseline scores were calculated from three formal exams before the fieldwork and converted to a scale of 10 to match up with other quizzes. In the following nine weeks, a quiz was designed each week. All the questions in the quizzes were selected from the teaching materials, therefore all the quizzes were closely related to the learning contents.

The achievement of all students during the fieldwork is illustrated in Figure 4.1 below:

Figure 4.1 Achievement of all the participants



As can be seen from Figure 4.1, the total achievement of all students across the four classes during the fieldwork showed a general downward trend in the first five weeks, then showed an upward trend for the remaining five weeks.

A paired t-test was conducted to compare the scores from every pair of consecutive weeks. The results in Table 4.2 indicate that only the downward changes between the second and third weeks, and then the fourth and fifth weeks, showed statistically significant differences. These two significant drops between Weeks 2 and 3 and then between Weeks 4 and 5 are clearly apparent from Figure 4.1.

Table 4.2 Comparison between every two weeks' score

	Paired Differences			
Week	Mean	t	df	Sig. (2-tailed)
1-2	-.12755	-1.690	195	.093
2-3	.14281	2.318	195	.021
3-4	-.05714	-.696	195	.487
4-5	.16449	2.504	195	.013
5-6	-.08878	-1.432	195	.154
6-7	.05316	.714	195	.476
7-8	-.07510	-.809	195	.420
8-9	.01806	.285	195	.776
9-10	-.05342	-.769	195	.443
10-1	.02347	.288	195	.774

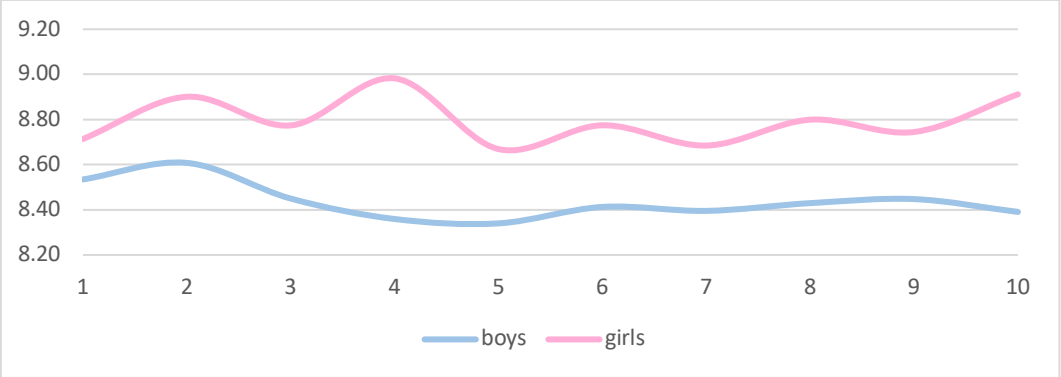
Differences in achievement between the two types of FA environment were also calculated and are depicted in Figure 4.2:

Figure 4.2 Achievement in different FA environments

In terms of FA environment, the scores from FA-rich (FAR) classes were higher than FA-lacking (FAL) classes in general. The final score from FA-rich classes was 9.07, which was slightly higher than the baseline (9.01). In the first three weeks, the FA-rich group's score dropped 2.33% from baseline, but increased by 2.49% in the last two weeks. Between the second week and the ninth week, the scores stayed around 8.86. On the other hand, in the FA-lacking environment, there was only a marginal difference between the baseline score (8.23) and the final score (8.21). It peaked at 8.64 in the second week and reached the bottom at 8.17 in the fifth week. It stayed consistent until the final week which showed a 1.44% decrease from 8.33 to 8.21. The average achievement score in the FA-rich environment was 8.89 and 8.32 in the FA-lacking environment. Generally speaking, achievement in the FA-rich environment is 6.41% higher than in the FA-lacking environment.

Figure 4.3 shows a comparison of boys' and girls' achievement:

Figure 4.3 Achievement between genders



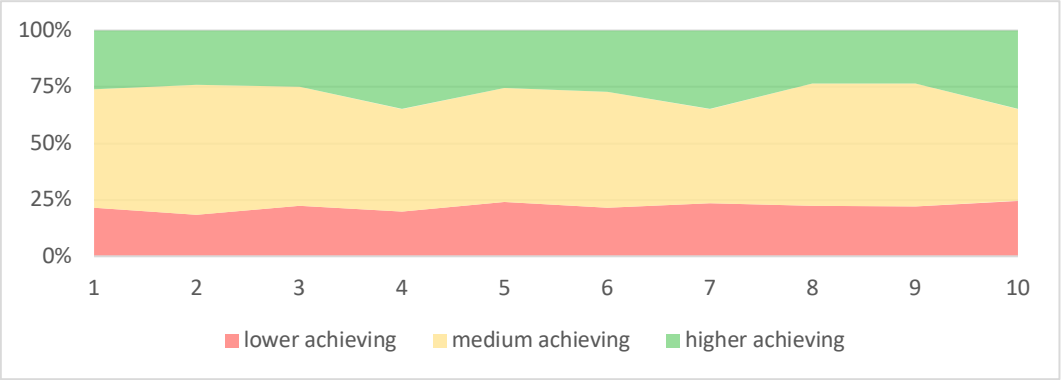
There was an impressive gap between boys' and girls' achievement, with girls performing better for the duration of the study.

Girls reached their highest score in the fourth week with 8.98 but fell to their floor score of 8.67 in the following week. Boys showed a broadly similar trend in the first three weeks, peaking at 8.61 in the second week, but reached the valley of 8.34 in the fifth week. After the fifth week, the girls' score climbed up gradually to 8.91 in the last week, which was 2.29% higher than the baseline; while boys' performance levelled off around 8.40 with an obvious fall in the last week (from 8.45 to 8.39), which is 1.70% lower than the baseline. Boys' average achievement score is 8.44, while for girls it is 8.80. Thus, girls' achievement is 4.10% higher than boys' achievement.

It was worth noting that all the discussion of gender difference was referred to the boys and girls in this study. As will be mentioned again in the discussion (e.g. the end of 6.2.1), the problematic concept of gender difference was not taken for granted in this study.

Other than the two variables discussed above, FA environment and gender, a third important variable requires consideration: achievement level. On the basis of the achievement scores, the learners were categorised into three groups: higher-achieving (HA), medium-achieving (MA) and lower-achieving (LA). Score cut-off points were set at 8.00 and 9.50. Students who achieved a score lower than 8.00 were categorised as lower-achieving; those who achieved a score higher than 9.50 were categorised as the higher-achieving group; the remaining students who scored between 8.00 and 9.50 were in the medium-achieving group. Figure 4.4 shows the percentage of the three achievement level groups in all the quizzes in the fieldwork.

Figure 4.4 Percentage of achievement groups



The percentage of the overall student cohort represented by the HA group stayed around 25% to 35%, the LA group stabilised around 20% to 25%. The distribution of the three groups was stable and reasonable. The figures show a satisfying result from the selected cut-off points.

4.4 Motivation

4.4.1 Quality of the questionnaire

Internal validity and reliability are used to verify the stability of an instrument across survey populations and they are central to experimental design, as specified in the Standards for Educational and Psychological Testing (Association, 1999).

Statistical tests of validity include conducting factor analyses and reliability tests include generating internal consistency estimates.

Reliability

In the main study, the questionnaire was distributed twice - once before and once after the fieldwork. The reliability analyses of the 20 motivation questions in the questionnaire produce Cronbach's alpha values: .901 in the pre-test and .882 in the post-test. Table 4.3 shows the Cronbach's Alpha values if any question is deleted. Every question held its value in the questionnaire, evidenced by the high value of the Cronbach's alpha for each.

Table 4.3 Reliability in main study-motivation questions

	Cronbach's Alpha if Item Deleted	
	Pre-test	Post-test
Motivation Question 1	.897	.875
Motivation Question 2	.896	.874
Motivation Question 3	.895	.873
Motivation Question 4	.895	.871
Motivation Question 5	.895	.877
Motivation Question 6	.895	.874
Motivation Question 7	.902	.881
Motivation Question 8	.894	.874
Motivation Question 9	.894	.872
Motivation Question 10	.896	.874
Motivation Question 11	.892	.871
Motivation Question 12	.892	.871
Motivation Question 13	.899	.882
Motivation Question 14	.895	.872
Motivation Question 15	.897	.877
Motivation Question 16	.908	.891
Motivation Question 17	.899	.881
Motivation Question 18	.900	.883
Motivation Question 19	.898	.884
Motivation Question 21	.897	.878

The questionnaire content was based on the seven motivation theories in language learning, which are: Value theory, Goal setting theory, Goal orientation theory, Attribution theory, Self-efficacy theory, Achievement theory and Extrinsic factors in motivation (Figure 2.7 Construct of Motivation Theories).

Table 4.4 sets out the Cronbach's Alpha coefficient for seven groups of questions, each one drawing on a specific motivation theory. It shows that all the alpha coefficients were higher than .70, confirming internal consistency of the questionnaire on motivation in language learning from the aspect of motivation theories.

Table 4.4 Reliability in main study-motivation theories

	Cronbach's Alpha if Item Deleted	
	Motivation	Motivation (II)
Value theory	.830	.787
Goal Setting theory	.831	.796
Goal Orientation theory	.840	.811
Attribution theory	.829	.788
Self-efficacy theory	.817	.780
Achievement theory	.841	.813
Extrinsic factors	.871	.853

Validity

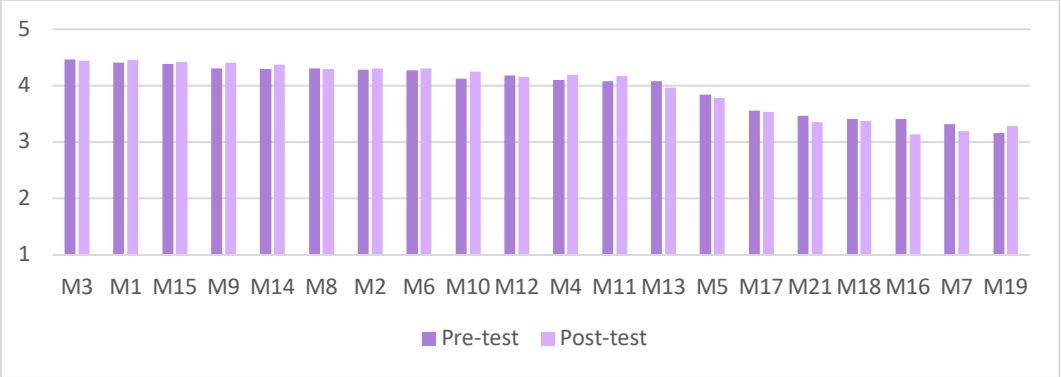
The validity of a survey instrument generally refers to the process of ensuring that the survey accurately measures what it is intended to measure. Internal validity explains how an experimental design is structured and encompasses all the steps of the scientific research method. With 196 survey respondents, the Pearson correlation values between each motivation question and motivation in general were at the significance level of .01. Each item was proven valid.

In summary, the questionnaire on learner motivation designed for this study had high validity and reliability which can be used to test students and generate data for analysis.

4.4.2 Factors in motivation

To understand the overall pattern in learners' responses to the questionnaire, the mean pre-test and post-test scores for each question were compared and are shown in Figure 4.5. There were 20 items in the questionnaire from M1 to M21 (M20 was deleted). It was apparent that M7, M16, M17, M18, M19, M21 generated lower scores in the questionnaire, which indicated that learners had lower motivation in performance goal, fear of failure and extrinsic factors. From the result of a paired-sample t-test (Appendix, Table 1), only the 16th question (M16): *I am afraid of the feeling of failure*, had a statistically significant decrease from the pre-test (M=3.41, SD=1.599) to the post-test (M=3.15, SD=1.558), t (192) =2.034, p<.05 (2-tailed). At the time of the post-test, being afraid of failure was not as important to learners as at the time of the pre-test.

Figure 4.5 Comparison of motivation questions between the pre-test and the post-test



It was also important to know which factors or aspects had greater impact on learners' motivation. Therefore, factor analysis was conducted to identify the most important motivation factors.

Factor analysis with motivation questions was run in SPSS to produce a Scree Plot which visually shows which components or motivation factors explain most of the variability in the data. Scree Plot is a means to determine the number of factors within a set of data. In the figure below, the line starts to show a mild trend from the third point, meaning each successive factor is accounting for smaller and smaller amounts of the total variance. The number of factors was determined by the number of points before the slope of the curve is clearly levelling off in Scree Plot of Factor Analysis in SPSS. Therefore, according to the Scree Plot below, there were three important factors in motivation on the basis of the data in motivation questionnaire.

Scree Plot in Factor Analysis

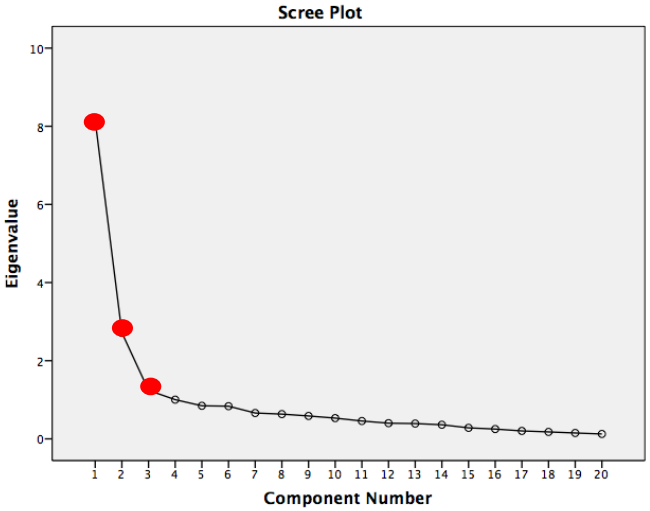


Table 2 Component Matrix in the Appendix illustrates the first results for motivation questions in different motivation factors. However, there were many motivation questions falling into two factors, which meant the factors had no clear boundaries. For example, both M5 and M6 belong to the first and the fourth factors. According to the Scree Plot, only three factors were considered. Therefore, to present a clearer factor-component relationship, Promax rotation method was used to adjust the results. Three factors with a range of motivation questions are presented clearly in Table 3 of the Appendix.

Different motivation questions fell into three motivation factors: *Language, Learner* and *Learning Environment*. These are highlighted in Table 3 of the Appendix. Meanwhile, since M21, M7, M13 and M16 had little impact on language learning motivation from the results of the factor analysis, they were not taken into consideration. Comparing to the motivation framework (Table 2.11), the grouping of motivation questions in each factor was not exactly the same as the motivation aspects, but was similar. Table 4.5 shows the relationship of aspects, theories, motivation questions, and factors. The three motivation aspects that were summarised from the literature review, were treated as the three main motivation factors in the following analyses: *Language*, *Learner*, *Learning Environment*. The most important factors relating to students' motivation related to *Language* itself. How students see themselves as language *Learners* was the second important factor. Finally, the *Learning environment*, represented by extrinsic reasons, was the third important factor. There are undoubtedly many other reasons that motivate students in language learning; these three factors were found in this study and will be analysed in the following sections.

Table 4.5 Factors in motivation framework

Aspects	Theory	Sub-dimension	Motivation Questions	Factors
Language	Value	Attainment value	M1	2
		Intrinsic value	M2	1
		Extrinsic value	M3	1
		Cost	M4	1
	Goal setting	Difficulty	M5	1
		Specificity	M6	1
Learner	Goal orientation	Performance goal	M7	
		Mastery goal	M8	2
	Attribution	Ability	M9	1
		Effort	M10	2
	Self-efficacy	Level	M11	2
		Strength	M12	2
	Achievement	Need for achievement	M13	
		Expectancy of success	M14	2
		Incentive value of success	M15	2
		Fear of failure	M16	
Learning environment	Extrinsic factors	Teacher	M17	3
		Parents	M18	3
		Peer	M19	3
		Education system	M20	
		Culture	M21	

Factor 1: *Language*
Factor 2: *Learner*
Factor 3: *Learning Environment*

4.4.3 Variation in motivation reported

The investigator observed the changes in students' motivation over 10 weeks of fieldwork. However, it seems that this period may not be long enough for significant changes in motivation to emerge. In this short time period, every minor change could seem like a signal for future improvement or drop. Figure 4.6 below depicts the changes in motivation overall and for each of the motivation factors *Language, Learner* and *Learning Environment*. As the figure shows, motivation levels did not change overall. Different motivation factors all

showed only marginal increases. Given that the questions identified as having little impact were discounted (see Figure 4.5), it was a reasonable expectation that the factors would show changes, but the totals stayed very similar.

Figure 4.6 Motivation in the pre-test and the post-test

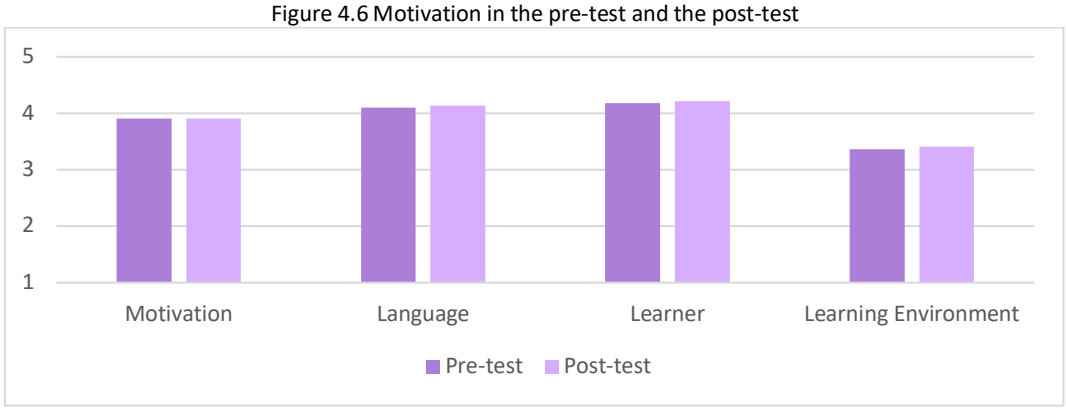


Figure 4.7 compares the average motivation levels of the LA, MA and HA groups in the pre-test and the post-test for each motivation factor and overall. For MA and HA, their motivation was mainly from *Language* and *Learner* factors. For LA, their motivation was fairly evenly spread across the three factors.

Figure 4.7 Motivation in different achievement groups

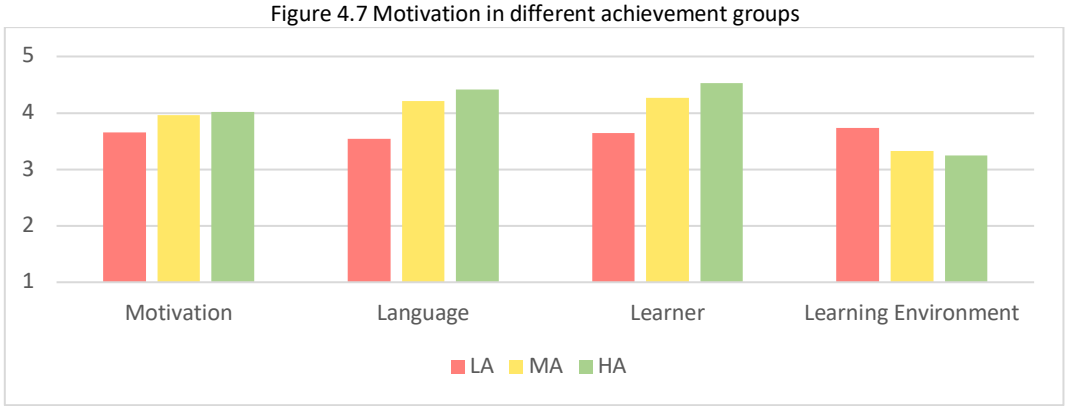


Figure 4.8 shows that students' overall motivation levels in the FA-rich (FAR) environment and the FA-lacking (FAL) environment were close to the same, but there are differences by motivation factor. FAR students paid more attention to *Language* than FAL, meanwhile, FAL were more motivated by *Learning Environment* than FAR. Students from both FA environments did not show much difference in motivation due to the factor of *Learner*.

Figure 4.8 Motivation in different FA environments

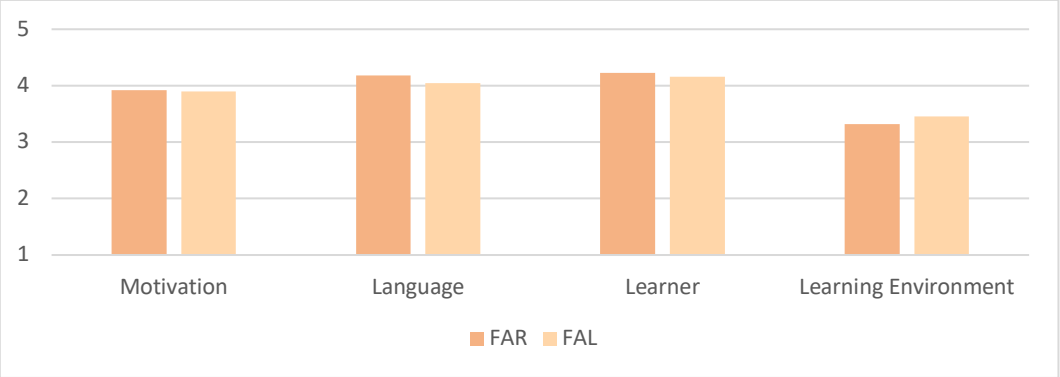


Figure 4.9 presents the average motivation levels of boys and girls in the pre-test and the post-test, grouped by motivation factor and overall motivation. In general, girls were more motivated than boys, but the differences were not significant. Boys' motivation was lower than girls' from the *Language* and *Learner* factors, and higher than girls' from *Learning Environment*.

Figure 4.9 Motivation in different genders

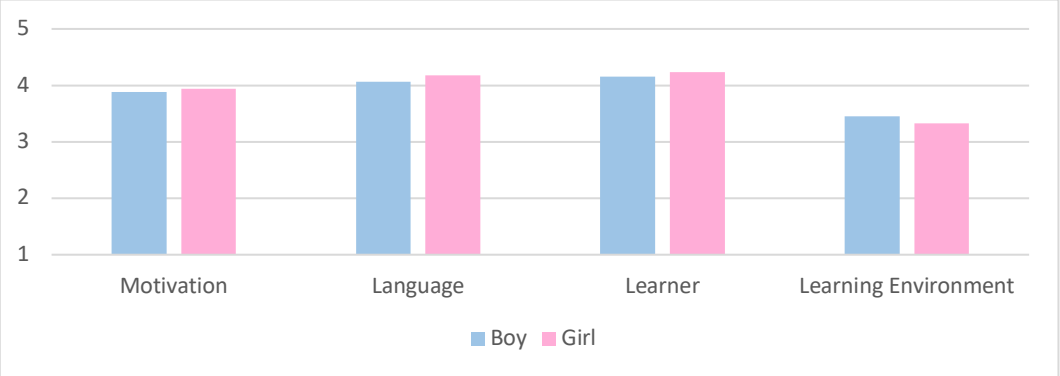


Figure 4.10 shows the changes in motivation in different contexts.

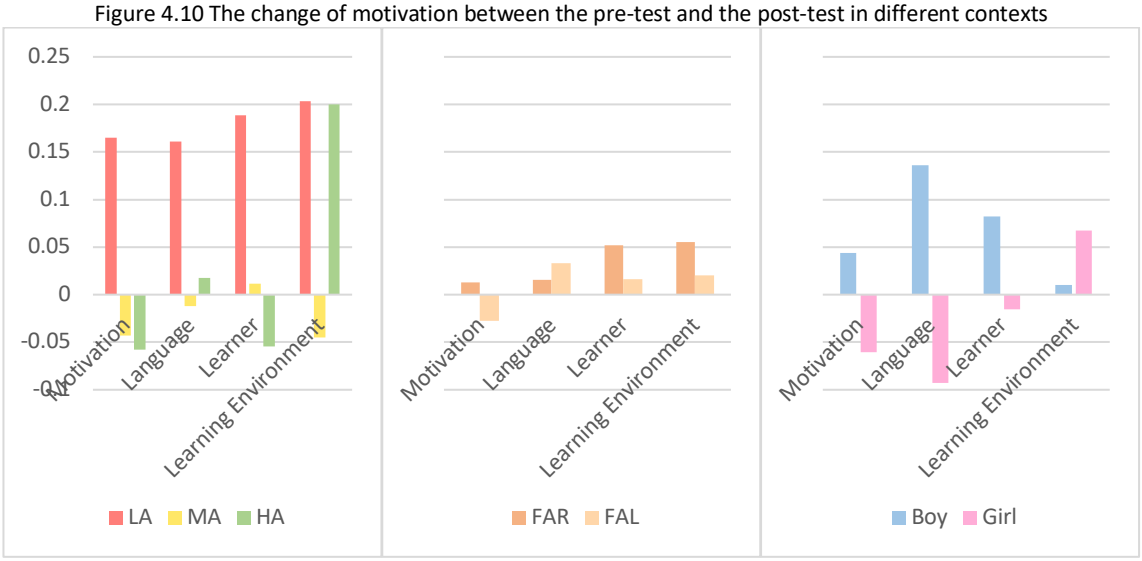
In terms of the different achievement groups, LA's motivation increased during the fieldwork, particularly in the factor of *Learning Environment*. Learning environment had a bigger and bigger impact on LA. MA's motivation showed a decrease, mainly in *Learning Environment*, but the change was very small. HA students had lower motivation in the post-test, but the factor of *Learning Environment* became more and more important. Like LA students, HA students paid more attention to *Learning Environment* as time went on.

In terms of FA environments, the changes were not significant. There was not much difference in motivation during the fieldwork in different FA environments. FAR's motivation increased while FAL's decreased. In each factor, FAR and FAL both showed increasing motivation. FAR had larger growth than FAL in *Learner* and *Learning Environment*[1].

[1] As shown in the discussion in 4.4.2, and Table 3 in Appendix, three questions were not included in motivation factors in this study, that may cause the decreasing of general motivation in FAL, but the increasing in each motivation factor.

With regard to gender, boys' motivation increased while girls' decreased. Boys started to pay more attention to *Language* and *Learner*. Girls' motivation in *Language* decreased, but they had more interest in *Learning Environment*.

Figure 4.10 The change of motivation between the pre-test and the post-test in different contexts



The figures above showed that a) the total motivation did not change during the fieldwork; b) *Language* and *Learner* were more important motivation factors than *Learning Environment*; c) the changes in motivation in different achievement groups and gender were more significant than the differences in FA environment.

4.5 Formative Assessment

Two surveys about FA were conducted along with the questionnaire about motivation before and after the fieldwork. The surveys explored the use of FA in the class and students' attitude towards FA strategies. This section shows the results from the surveys. In the following figures and tables, the names of FA strategies are abbreviated as following:

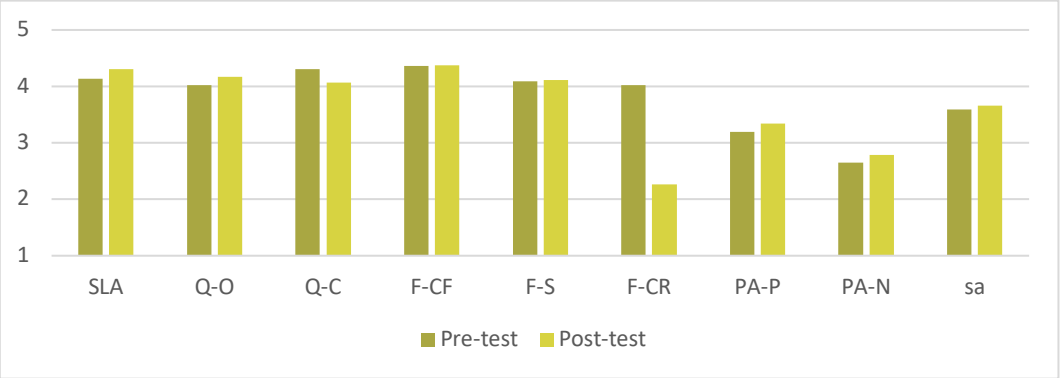
Sharing learning aims	SLA
Open-ended questioning	Q-O
Close-ended questioning	Q-C
Confirmative feedback	F-CF
Suggestive feedback	F-S
Corrective feedback	F-CR
Positive peer-assessment	PA-P
Negative peer-assessment	PA-N
Self-assessment	sa (distinguish from SA: summative assessment)

4.5.1 Use of FA

Figure 4.11 shows the use of FA from students' perspective in the pre-test and the post-test. Confirmative feedback was the most popular FA strategy, while negative peer-assessment was the least. The use of sharing learning aims, questioning and feedback were more

frequent than peer-assessment and self-assessment. After the 10 weeks of fieldwork, the use of close-ended questions and corrective feedback decreased, and the corrective feedback dropped to the lowest point. Other strategies were all slightly increased. As time went on, teachers used much less corrective feedback than before.

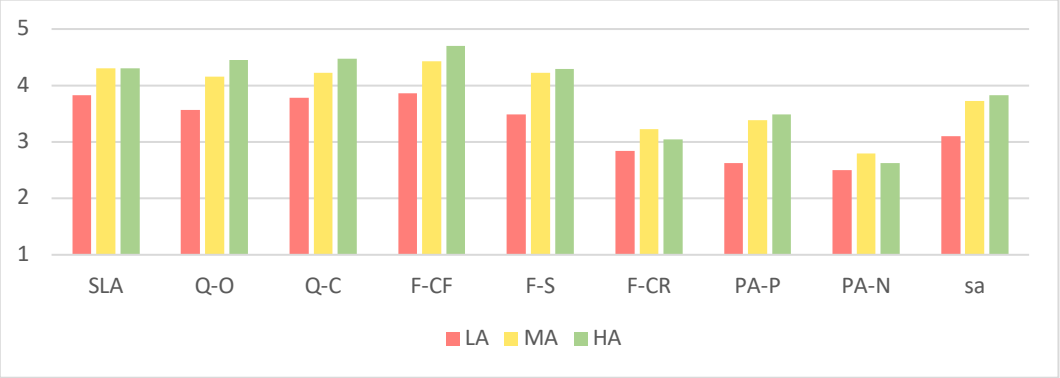
Figure 4.11 Use of FA in the pre-test and the post-test



To explore the change in the use of FA in different contexts, the importance of each strategy was compared for different sub-groups. Figure 4.12, Figure 4.13 and Figure 4.14 set out the FA used by different achievement groups, different FA environments and different genders.

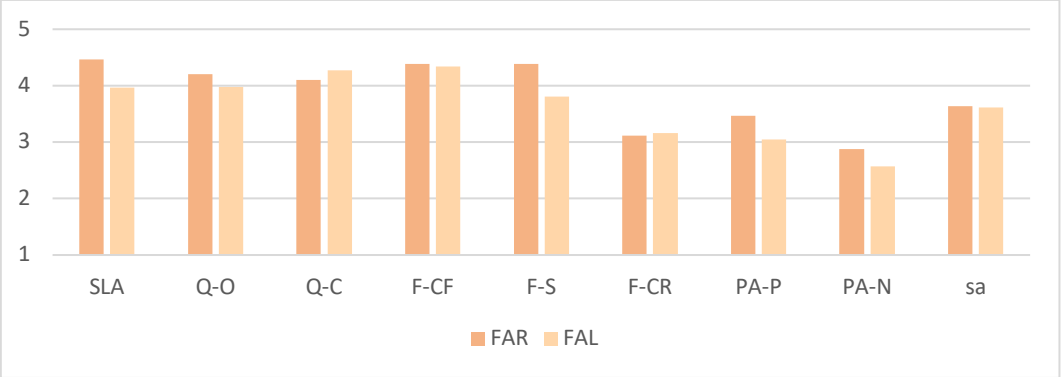
In terms of achievement level groups, LA students experienced the least amount of FA in the class. In general, the lower the scores that students achieved, the less FA they received in the class. Henly (2003) also found that top ten percent of the class accessed FA twice as often as students in the bottom ten percent of the class. However, MA students showed more sharing of learning aims, corrective feedback and negative peer-assessment than HA and LA students.

Figure 4.12 FA in different achievement groups



In terms of FA environments, students in FAR experienced more FA than those in FAL. However, FAL students used more close-ended questioning and corrective feedback than FAR students. These two environments showed bigger differences in sharing learning aims, suggestive feedback and positive peer-assessment.

Figure 4.13 FA in different FA environments



With regard to gender, girls received slightly more FA than boys in general, and boys used more negative peer assessment, but none of the gender differences were statistically significant.

Figure 4.14 FA in different genders

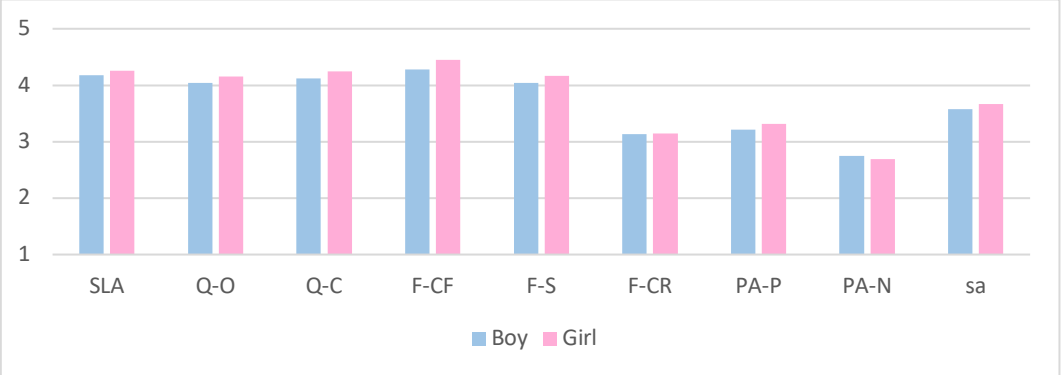


Figure 4.15 shows the changes of FA between the pre-test and the post-test in different contexts. The most stable FA strategy was confirmative feedback. There was little change in it between the pre-test and the post-test. It was obvious that the use of corrective feedback dropped dramatically, especially in the HA group. This was a reasonable expectation because the higher the students' achievement, the fewer mistakes they would make, therefore, the teacher would not need to correct them. However, it can be seen from the middle chart in Figure 4.15 that both FA environments had less corrective feedback, so one of the reasons that students noticed the lesser use of corrective feedback might be the change in the teachers' behaviour. It is very clear that by the end of the fieldwork students knew more about FA than at the start. Their perceptions may affect their responses too. Apart from corrective feedback, the use of other FA strategies changed in the following contexts.

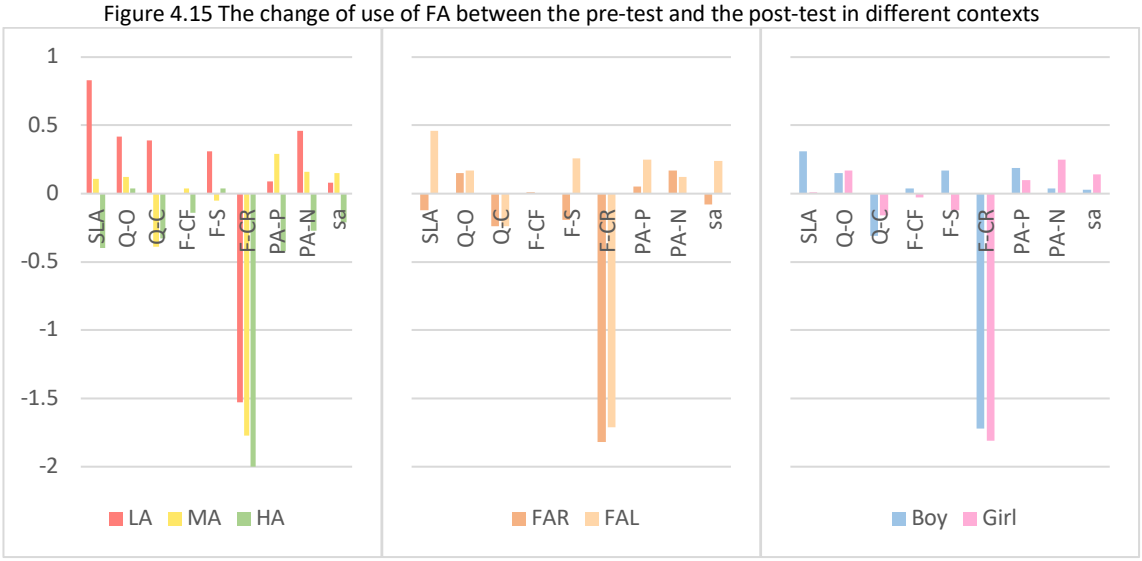
In terms of achievement levels, LA students noticed more FA strategies had been used with them, sharing learning aims in particular. Smaller amounts of corrective feedback, close-ended questioning and suggestive feedback were used on MA students. MA experienced more FA strategies of other kinds. Less FA was used with HA students, with the exception of open-ended questioning and suggestive feedback.

In terms of FA environments, FAL students saw more use of suggestive feedback and self-assessment, and especially sharing learning aims. FAR students experienced fewer of these

strategies. Students in both environments used more open-ended questioning, peer-assessment and less close-ended questioning.

With regard to gender, the differences were even smaller. Boys noticed that more FA was used, especially sharing learning aims, but less close-ended questioning was used. For girls, negative peer-assessment was the strategy that they used more often.

Figure 4.15 The change of use of FA between the pre-test and the post-test in different contexts

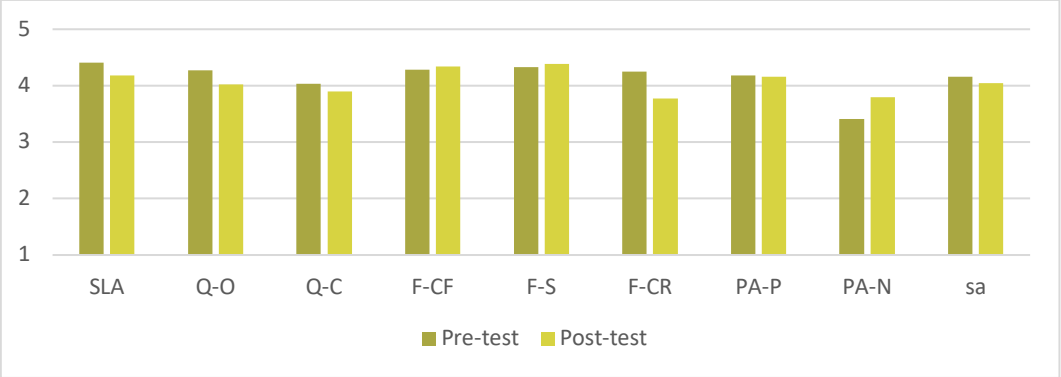


The figures above showed that: a) the FA strategies employed by teachers were more often used than those initiated by students; b) the use of corrective feedback had shrunk dramatically; but the use of confirmative feedback was the steadiest; c) the group differences between achievement levels were more significant than those between FA environment or genders.

4.5.2 Attitude to FA

Figure 4.16 shows the attitudes to FA from the students' perspective in the pre-test and post-test. Students liked the suggestive feedback the most, and negative peer-assessment the least. They liked confirmative feedback, suggestive feedback, negative peer-assessment more after 10 weeks, but not the other strategies. Students' attitudes towards corrective feedback dropped the most, and their attitudes towards negative peer-assessment increased the most among all the FA strategies.

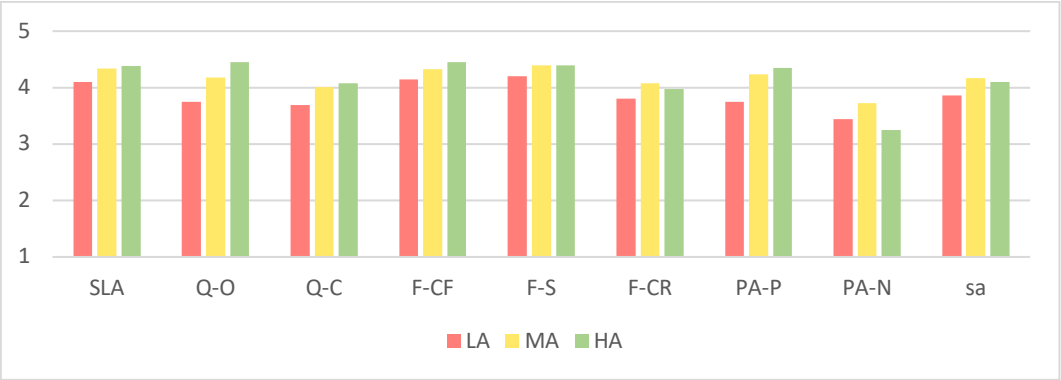
Figure 4.16 Attitude to FA in the pre-test and the post-test



The changes in the attitude to FA in different contexts are shown in the figures below. Figure 4.17, Figure 4.18 and Figure 4.19 depict the attitudes to FA in different achievement groups, different FA environments and different genders.

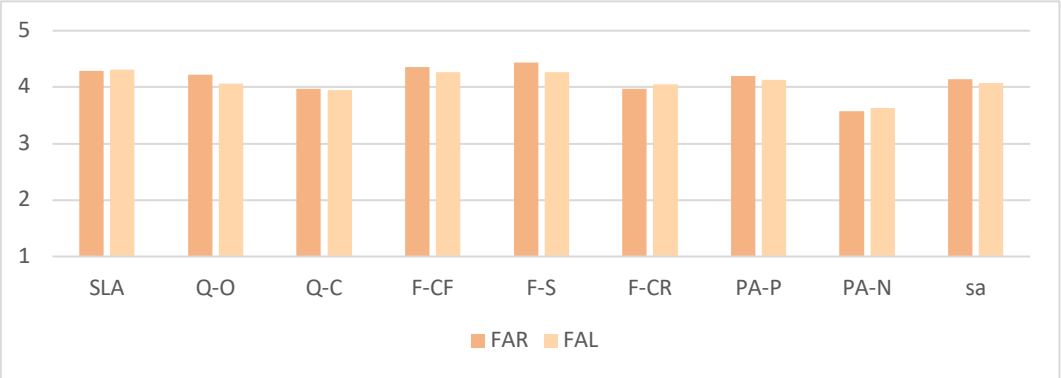
In terms of achievement levels, for most of the strategies, the higher the students' achievement, the more they like the strategy. However, this was not the picture for corrective feedback, negative peer-assessment and self-assessment. The most significant difference was that HA students did not like negative peer-assessment.

Figure 4.17 Attitude to FA in different achievement groups



In terms of FA environments, the difference between the two groups was not great. The biggest difference lies in suggestive feedback. FAL students showed more interest in sharing learning aims, corrective feedback and negative peer-assessment than FAR students.

Figure 4.18 Attitude to FA in different FA environments



With regard to gender, girls tended to like all the FA strategies more than the boys, as apparent from Figure 4.19.

Figure 4.19 Attitude to FA in different genders

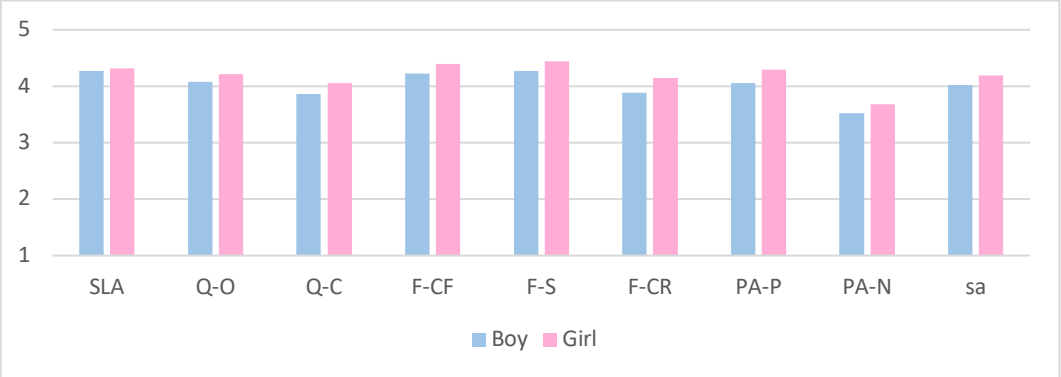
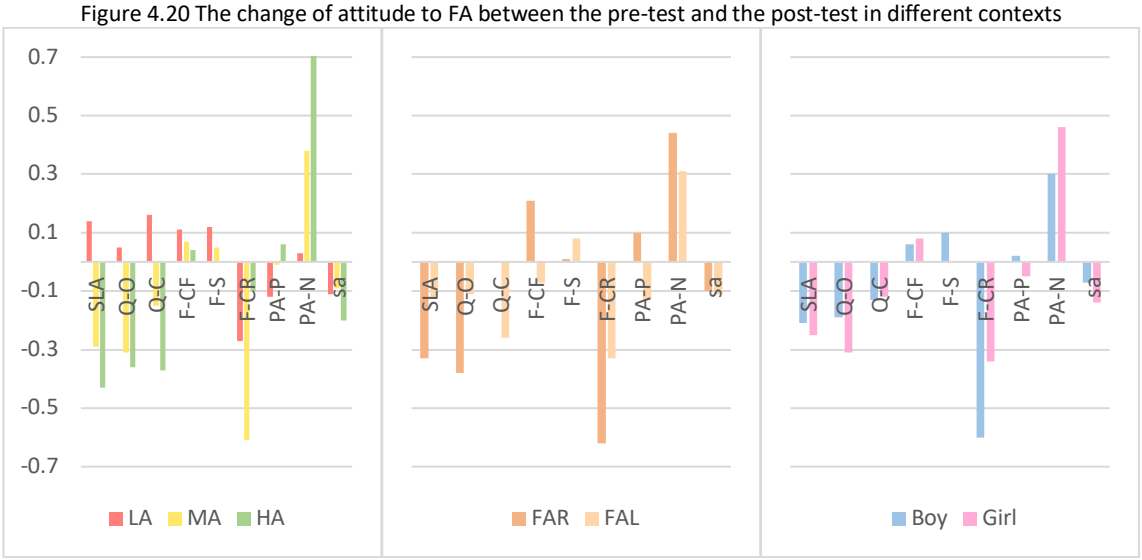


Figure 4.20 illustrates the changes in attitude to FA in the contexts of FA environment, achievement levels and gender. In general, the trend of students' attitude changes was similar. Two obvious changes were corrective feedback and negative feedback. Students liked corrective feedback less but grew more interest in negative peer-assessment. Attitudes to confirmative feedback, suggestive feedback, positive peer-assessment and self-assessment were comparatively stable.

Students showed less interest in corrective feedback and self-assessment, but more in confirmative feedback, suggestive feedback and negative peer-assessment. Students' attitudes to sharing learning aims and questioning were decreasing in every group except LA students. Attitudes to positive peer-assessment varied in different groups. LA students, FAL students and girls showed less interest in it, but HA students, FAR students and boys liked this form of FA more.

Figure 4.20 The change of attitude to FA between the pre-test and the post-test in different contexts



The figures above showed that: a) students like suggestive feedback the most, and negative peer-assessment the least; b) their attitudes to FA are positively correlated with achievement levels, but not in the case of corrective feedback, negative peer-assessment

and self-assessment; c) girls liked FA more than boys; d) students' interests decreased in corrective feedback and increased in negative peer-assessment.

4.6 Relationship between FA and motivation

To answer the research question '*Is there a relationship between FA and students' motivation?*', the correlations between FA and motivation were calculated in SPSS.

4.6.1 In general

Correlations between FA and motivation were mainly moderate in the pre-test; and were small in most cases in the post-test (Table 4.6). There were two exceptions, negative peer-assessment in the pre-test and self-assessment in the post-test. Negative peer-assessment had no correlation with motivation in the pre-test. In the post-test, self-assessment had a significant relationship with motivation. Compared to student-led FA, such as peer-assessment and self-assessment, FA from teachers, which included sharing learning aims, questioning and feedback, had weaker relationships with motivation all the time.

In the pre-test

Confirmative feedback had a small correlation with motivation ($r=.249$, $p<.01$); the rest of the FA strategies had relationships with motivation of medium strength except for negative peer-assessment. In terms of motivation factors, there was no FA strategy which had a significant relationship with *Learning Environment*. Every strategy apart from negative peer-assessment had moderate correlations with the other two factors *Language* and *Learner*. Sharing learning aims ($r=.513$, $p<.01$) was strongly related to motivation with respect to *Language*.

In the post-test

The strong relationships between sharing learning aims and *Language* became weak. It was notable that self-assessment had large correlations with *Language* ($r=.512$, $p<.01$) and *Learner* ($r=.529$, $p<.01$), as well as motivation in general ($r=.560$, $p<.01$). More FA strategies had correlations with *Learning Environment* in the post-test, however, the strength of the relationships was still weak.

After the 10-week fieldwork period, the relationships motivation had with sharing learning aims, open-ended questioning, corrective feedback were weakened, but the rest of the FA strategies became more and more important for motivation, particularly self-assessment.

With regard to motivation factors, *Learning Environment* had the weakest relationship with FA, meanwhile *Language* had the strongest relationship with FA strategies.

In terms of FA strategies, self-assessment played the biggest role, while negative peer-assessment had the weakest relationship with motivation.

Table 4.6 Correlation between FA and motivation

		Pre-test				Post-test			
		Language	Learner	Learning Environment	Motivation	Language	Learner	Learning Environment	Motivation
SLA	r	**.513****	.477**	.018	.451**	.297**	.267**	.136	.352**
	p	.000	.000	.807	.000	.000	.000	.063	.000
Q-O	r	.438**	.451**	.019	.369**	.433**	.339**	.045	.364**
	p	.000	.000	.803	.000	.000	.000	.542	.000
Q-C	r	.386**	.406**	.005	.327**	.354**	.293**	.108	.347**
	p	.000	.000	.943	.000	.000	.000	.141	.000
F-CF	r	.330**	.333**	.027	.249**	.431**	.329**	.187*	.346**
	p	.000	.000	.717	.001	.000	.000	.010	.000
F-S	r	.470**	.477**	.006	.391**	.435**	.362**	.170*	.416**
	p	.000	.000	.934	.000	.000	.000	.020	.000
F-CR	r	.370**	.406**	.121	.383**	.148*	.159*	.205*	.273**
	p	.000	.000	.104	.000	.042	.029	.005	.000
PA-P	r	.474**	.443**	.098	.389**	.486**	.437**	.208*	.482**
	p	.000	.000	.186	.000	.000	.000	.004	.000
PA-N	r	-.099	-.099	.128	-.005	.107	.134	.229*	.264**
	p	.186	.183	.084	.946	.143	.066	.002	.000
sa	r	.446**	.391**	.139	.381**	**.512****	**.529****	.253**	**.560****
	p	.000	.000	.061	.000	.000	.000	.000	.000

** Correlation is significant at the .01 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

r= Pearson Correlation

p= Sig. (2-tailed)

4.6.2 Achievement group

In different achievement groups, the correlations showed varied patterns. Table 4 in the Appendix illustrates the strength of the correlations. The following paragraphs discuss the findings from Table 4.

Lower-achieving group (LA)

In the pre-test, the relationship between FA and motivation was very close and strong. Positive peer-assessment (r=.699, p<.001), sharing learning aims (r=.691, p<.001) and suggestive feedback (r=.668, p<.001) had the strongest correlations with motivation. Open-ended questioning (r=.602, p<.001), close-ended questioning (r=.539, p<.01) and corrective feedback (r=.526, p<.01) also had strong relationships with motivation. Confirmative feedback and self-assessment were related to motivation too, but with a relatively moderate strength. Negative peer-assessment (r=-.416, p<.05) had a moderate but negative relationship with lower achievers' motivation.

After 10 weeks, open-ended questioning and corrective feedback lost their strong connection to motivation. Other FA strategies had smaller Pearson Correlation values

except for confirmative feedback, negative peer-assessment and self-assessment. Self-assessment's Pearson Correlation value was boosted from .369 to .637, to become the strongest relationship in the post-test. Negative peer-assessment changed greatly from a negative to a moderately positive relationship (r=.483, p<.01).

With regard to motivation factors, *Learning Environment* yielded only three relationships during the fieldwork: with sharing learning aims (r=.355, p<.05), positive peer-assessment (r=.596, p<.001) and self-assessment (r=.536, p=.001), all in the post-test. The rest of the FA strategies had no relationship with *Learning Environment*. Motivation in *Language* and *Learner* had positive correlations with all FA strategies in the pre-test, strong or moderate, excepting negative peer-assessment which had negative moderate relationship with these two factors. Most of the FA strategies had weaker relationships with *Language* and *Learner* in the post-test*,* especially the corrective feedback which had no relation with *Language* and *Learner* in the post-test*;* only negative peer-assessment and self-assessment developed stronger relationships with motivation factors.

Medium-achieving group (MA)

Most of the correlations for MA were significant and moderate. In the pre-test, the strongest relationship between FA and motivation was with corrective feedback (r=.307, p=.001), and in the post-test, it was self-assessment (r=.448, p<.001). In general, the relationships became stronger in the post-test, but they were still moderate or weak. In the post-test, every FA strategy had a significant relationship with motivation. In terms of FA strategies, it was self-assessment that had the strongest relationship with motivation and its factors (*Language* and *Learner*). Similarly to the lower-achieving group, medium achievers did not show much correlation between FA and *Learning Environment*.

Higher-achieving group (HA)

Compared to the pre-test, FA yielded stronger and more relationships with motivation in the post-test for HA students. The medium-strength correlation with sharing learning aims (r=.409, p<.05) in the pre-test disappeared in the post-test; but more medium-strength correlations appeared in the post-test, such as close-ended questioning (r=.442, p<.05), confirmative feedback (r=.432, p<.05) and suggestive feedback (r=.408, p<.05). Corrective feedback kept its moderate correlation in the pre-test (r=.422, p<.05) and the post-test (r=.407, p<.05); open-ended questioning and negative peer-assessment produced no relationship with motivation in either test. Positive peer-assessment and self-assessment were highly correlated to motivation, and the strength increased in the post-test. Self-assessment ($r=.701^{\text{pre-test}}/.808^{\text{post-test}}$, p<.001) had the strongest relationship with motivation among all the FA strategies.

In terms of motivation factors, HA did not show any correlations between FA and *Learning Environment* in the pre-test*,* except for self-assessment (r=.442, p<.05). Compared to the other two achievement groups, more FA strategies were correlated to *Learning Environment* and fewer FA strategies were related to *Learner* in the HA group in the post-test. Only two strategies - positive peer-assessment (r=.590, p=.001) and self-assessment (r=.661, p<.001) - were related to motivation in *Learner* in the post-test. Six out of nine FA

strategies were either strongly or moderately related to *Learning Environment*, and these were sharing learning aims (r=.417, p<.05), close-ended questioning (r=.474, p<.05), confirmative feedback (r=.514, p<.01), positive peer-assessment (r=.632, p<.001), negative peer-assessment (r=.492, p<.01) and self-assessment (r=.673, p<.001). Meanwhile, the other two achievement groups had only one or two significant correlations between FA and *Learning Environment*.

To summarise the correlations between FA and motivation for the three groups, all the information above was compared in terms of correlation features, pre-test/post-test, FA strategies and motivation factors.

1) The strength of the correlations in HA and LA was strong, but it was comparatively weak in MA. The achievement group which FA had the greatest impact on was LA. The trend was: the higher the achievement, the less the impact of FA on students' motivation. However, in the previous sections, the findings showed that the more students achieved, the stronger their motivation became (Figure 4.7), the more FA was used on them (Figure 4.12) and the more students liked FA (Figure 4.17). Therefore, HA students' motivation may derive from other factors, not just the FA strategies examined in this research. In contrast, LA students' motivation, even though it was not as strong as the HA group's, had strong correlations with FA. In other words, FA's impact was much bigger in LA than in HA.
2) The relationship between FA and motivation would become weaker after a period of study time in LA, but stronger in HA. Nevertheless, there were exceptions. In the post-test, negative peer-assessment and self-assessment had stronger correlations with motivation in LA; sharing learning aims and corrective feedback had weaker correlations with motivation in HA. FA need to be maintained carefully by teachers and LA students. Even very effective strategies, as evidenced by the strong correlations in LA in the pre-test, can lose their impact on students' motivation without good maintenance.
3) Of all the FA strategies, positive peer-assessment and self-assessment had much stronger relationships with motivation, particularly in HA. These two strategies played an important role in enhancing motivation, but were not the most often used FA strategies (Figure 4.11). This suggests a strong need to focus more on the most relevant, proven practices in the future. Also, the influence of negative peer-assessment turned from negative to positive in LA. The students had apparently started to accept the critical assessment from their peers.
4) In terms of motivation factors, *Language* had the strongest relationship with FA and *Learning Environment* had the weakest. Therefore, it would be more useful if teachers used FA strategies to promote students' motivation from the aspects of *Language* itself, as the impact of FA was bigger for *Language* than the other two factors. *Learner* was important in motivation in the beginning, but it had fewer correlations with FA later. LA students paid more attention to *Learner*. For LA students, it would be important to encourage them with aspects of *Learner*, to build LA students' self-esteem. By doing so, FA could stay effective in LA. Meanwhile, HA students paid more attention than the other two groups to *Learning Environment* aspects, especially in the post-test. *Learning Environment* was more important in the HA group than in other two groups. Teachers can use FA strategies related to *Learning Environment* to promote motivation in HA students.

4.6.3 FA environment

Table 5 in the Appendix shows the correlations between FA and motivation in FAR and FAL environments.

FA-rich environment (FAR)

Each FA strategy was significantly correlated to motivation, excepting negative peer-assessment in the pre-test. Most of the relationships were moderate, with a few in the post-test that were strong: suggestive feedback (r=.502, p<.001), positive peer-assessment (r=.516, p<.001) and self-assessment (r=.543, p<.001). All of the FA strategies had stronger relationships with motivation in the post-test than in the pre-test. In terms of motivation factors, *Language* and *Learner* were closely correlated with FA in the pre-test, while *Learning Environment* had no relationship with any of the strategies. In the post-test, open-ended questioning (r=.502, p<.001), confirmative feedback (r=.594, p<.001), and suggestive feedback (r=.515, p<.001) turned from moderately correlated into strongly correlated with *Language*, but meanwhile, corrective feedback and negative peer-assessment lost their connection with *Language* and *Learner*. *Learning Environment,* in contrast, had six new weak or moderate relationships with FA in the post-test. Noticeably, negative peer-assessment was negatively correlated with *Language* (r=-.239, p<.05) and *Learner* (r=-.221, p<.05) in the pre-test, and positively correlated with *Learning Environment* in the post-test (r=.278, p<.01). Of all the FA strategies, self-assessment had the strongest impact on FAR's motivation.

FA-lacking environment (FAL)

All the FA strategies were significantly related to motivation, except for negative peer-assessment in the pre-test. All the relationships were moderate or weak, apart from sharing learning aims in the pre-test (r=.525, p<.001) and self-assessment in the post-test (r=.555, p<.001). The strength of the correlations in the post-test in FAL was weaker than in FAR, with one exception: self-assessment. In terms of motivation factors, like FAR, *Language* and *Learner* were related to FA, while *Learning Environment* in the post-test only had one weak relationship with corrective feedback (r=.298, p<.01). Three FA strategies had significant and strong correlations with *Language* in the pre-test: sharing learning aims (r=.537, p<.001), open-ended questioning (r=.509, p<.001) and suggestive feedback (r=.550, p<.001). These strategies were also strongly related to *Learner*. However, in the post-test, only self-assessment had strong correlation, not only with *Language* (r=.636, p<.001), but also *Learner* (r=.572, p<.001).

To summarise the correlations between FA and motivation in two different FA environments, all the data above were compared in terms of correlation features, pre-test/post-test, FA strategies and motivation factors.

1) Generally speaking, FAR yielded stronger FA-motivation relationships than FAL. According to the figures before, FAR students had similar motivation with FAL (Figure 4.8), but more FA was used on them (Figure 4.13) and they had a more positive

attitude towards FA (Figure 4.18). The evidence chimes that the impact of FA was bigger in FAR.

2) Motivation and FA had stronger and more correlations in the post-test for FAR, especially in *Learning Environment*. For FAL, positive FA-motivation relationships declined for teacher-led FA strategies but increased for those related to students. FAR students expressed more and more motivation, especially from the external environment, along the time of the study; while FAL students felt that student-led assessment, such as peer-assessment and self-assessment, could motivate them better.
3) In FAR, the most influential FA strategies were positive peer-assessment in the pre-test, and suggestive feedback, positive peer-assessment, and self-assessment in the post-test. In FAL, the most influential FA strategies were sharing learning aims in the pre-test, and self-assessment in the post-test. Both FAR and FAL students paid more attention to student-led FA in the post-test, where even negative peer-assessment had a stronger relationship with motivation. Self-assessment played a vital role in both FA environments.
4) *Learning Environment* was not related to FA in the pre-test, but had many significant but weak correlations with FA in the post-test for FAR. During the study, external environment played a more and more influential role in FAR. The teacher provided abundant FA for students in FAR. The increasing impact of *Learning Environment* on motivation demonstrated the positive effects of FA. FA was significantly related to the other two factors: *Language* and *Learner*. Compared to *Learner*, *Language* had a stronger relationship with FA. The impact of FA on *Language* was stronger than on other motivation factors.

4.6.4 Gender

The correlations between motivation and FA in boys (B) and girls (G) are set out in Table 6 in the Appendix.

Boys

Boys had significant correlations between FA and motivation, but not between negative peer-assessment and motivation in the pre-test. However, most of the relationships were moderate. There was one exception: self-assessment in the post-test ($r=.604$, $p<.001$). The strength of the correlations between teacher FA was weaker in the post-test, excepting confirmative feedback ($r=.328^{\text{pre-test}}/.430^{\text{post-test}}$, $p<.001$); the strength of the correlations between student-led FA was stronger. Boys were more motivated by student assessment as the study proceeded. In terms of FA strategies, sharing learning aims, corrective feedback and negative peer-assessment had less connection with motivation factors than other FA strategies. These three strategies played a less important role in boys' motivation. On the other hand, self-assessment was the most influential strategy for boys after a period of study time, even though it had the weakest relation to motivation in the pre-test. With respect to motivation factors, FA, especially teacher FA strategies, were strongly related to *Language* and *Learner*, but not with *Learning Environment* in the pre-test. However, in the post-test, the strength of correlations between FA and *Language*/*Learner* were weakened (apart from self-assessment), the number of significant correlations was fewer. Three new

but weak correlations appeared between FA and *Learning Environment* in the post-test. Therefore, FA had less impact on boys' motivation factors of *Language* and *Learner,* but more impact on *Learning Environment* in the post-test than it had in the pre-test.

Girls

Six out of nine strategies were significantly related to girls' motivation in the pre-test. Most of the correlations were moderate, except for self-assessment ($r=.513^{\text{pre-test}}/.514^{\text{post-test}}$, $p<.001$) and positive peer-assessment in the post-test ($r=.547$, $p<.001$) which were strong. There were more and stronger correlations between motivation and FA for girls in the post-test. Even so, negative peer-assessment did not have any significant relationship with girls' motivation. Other FA strategies had weak or moderate correlations with motivation. In terms of motivation factors, FA had more correlations with *Language* and *Learner* than *Learning Environment*. The changes in the correlations before and after the fieldwork were not significant. Most of them were weak or moderate, besides the negative peer-assessment mentioned earlier, as well as self-assessment which had a strong relationship with *Language* ($r=.576^{\text{pre-test}}$, $p<.001$) and *Learner* ($r=.532^{\text{pre-test}}/.522^{\text{post-test}}$, $p<.001$); as did positive peer-assessment with Learner ($r=.511^{\text{post-test}}$, $p<.001$). *Learning Environment* had a few weak correlations with positive peer-assessment, negative peer-assessment and self-assessment in the pre-test, and also with positive peer-assessment and corrective feedback in the post-test.

In terms of correlation features, pre-test/post-test, FA strategies and motivation factors, the correlations of FA and motivation in different genders are summarised below:

1) Most of the correlations were moderate or weak; the only strong correlations were between motivation and positive peer-assessment and self-assessment respectively. Motivation (Figure 4.9) and attitude to FA (Figure 4.19) were slightly weaker in boys than girls. Less FA was used with boys (Figure 4.14). However, there was no significant difference in FA-motivation correlations between boys and girls.
2) Boys' FA-motivation correlations became weaker after the period of study, for teacher FA strategies in particular. Boys were not affected by self-assessment at first; interestingly, it became significantly important to them in the post-test. Meanwhile, the impact of other strategies on their motivation declined. Girls' FA-motivation correlations had increased slightly after the 10-week study, except for corrective feedback. The biggest change for girls was the increasing impact of positive peer-assessment.
3) Self-assessment was the most important FA strategy to boys and girls. Although boys thought that other strategies could motivate them more in the first place, they later found themselves more easily motivated by self-assessment. Girls thought student-led assessments, especially self-assessment, were important to them before and after the fieldwork. Although both boys and girls preferred student assessment, the different, changing routes of their focus was interesting.
4) Both boys and girls were more likely to be motivated by *Language* and *Learner*. There were no big differences between these two motivation factors. However, it was notable that in boys FA showed strong impacts on these two motivation factors in the pre-test, but not in the post-test. All the students were unlikely to be

motivated by FA in terms of *Learning Environment*, however FA began to have more impact on *Learning Environment* for boys in the post-test.

4.6.5 Matrix comparison of Achievement group and Gender

After analysing the interrelationship in different contexts separately, the following section will discuss the correlations between gender and achievement levels (set out in Table 7 in the Appendix). Each group had only around 35 students, therefore it is hard to generalise the findings to a broader field of students, but it would be helpful to understand the patterns of findings for the students in this study.

Unlike HA and LA groups, the MA group had weak or moderate correlations with only one exception: self-assessment with *Language* in girls' MA group ($r=.553$, $p<.001$). Between the pre-test and the post-test, there were more and stronger colorations between FA and motivation. FA was impacting MA students' motivation all the time, but not in a statistically significant way. Self-assessment was the most influential strategy in MA (also evidenced with Table 5 in the Appendix). Boys in this group started to realise the importance of self-assessment in the post-test ($r=.493$, $p<.001$); meanwhile girls knew its importance in the pre-test ($r=.492$, $p<.001$) and in the post-test ($r=.416$, $p=.001$). Besides this, the female MA group was also influenced by positive peer-assessment in the pre-test ($r=.342$, $p<.01$) and in the post-test ($r=.456$, $p<.001$). All the relatively strong correlations have been presented in this paragraph. Therefore, the other findings for the MA group will not be elaborated in the following section due to the fact that the correlations were not strong enough.

Boys' LA group

In this group, most of the correlations were moderate and strong. Students' motivation was strongly related to sharing learning aims ($r=.566$, $p<.05$), open-ended questioning ($r=.555$, $p<.05$), suggestive feedback ($r=.599$, $p<.01$) and positive peer-assessment ($r=.661$, $p<.01$) in the pre-test, and positive peer-assessment ($r=.649$, $p=.001$), negative peer-assessment ($r=.502$, $p<.05$) and self-assessment ($r=.754$, $p<.001$) in the post test. Their motivation influences shifted from teacher assessment to student assessment during the fieldwork.

In terms of FA strategies, close-ended questioning, confirmative feedback, negative peer-assessment and self-assessment were the strategies that had increasing impact on male LA students' motivation including each factor. Other strategies showed different trends. It was noticeable that the relationship between negative peer-assessment and motivation changed from negative to positive, while the positive peer-assessment had a slightly weaker relationship in the post-test than in the pre-test. The strong and significant correlation between self-assessment and motivation made it the most influential FA strategy in the post-test. Students in the male LA group had started to understand the positive impact of critical peer-assessment and self-assessment.

In terms of motivation factors, FA strategies had more impact on *Learner* aspects for male LA students in the pre-test and on *Language* aspects in the post-test. For this group, FA's impact was bigger on *Language* as the study progressed, and smaller on *Learner*. *Learning Environment* had a negative correlation with negative peer-assessment in the pre-test, and

was significantly and strongly related to positive peer-assessment and self-assessment in the post-test. The negative impact on *Learning Environment* disappeared and was replaced by a more positive and strong correlation.

Boys in the LA group were highly motivated by sharing learning aims, open-ended questioning, suggestive feedback and positive peer-assessment in the beginning; later, they were mainly motivated by positive peer-assessment and self-assessment. The most important motivation factor was *Learner* in the pre-test, and *Language* in the post-test. Student assessment had more and more impact on *Learning Environment* for the male LA group.

Boys' HA group

This was the group that had least correlations between FA and motivation. Teacher FA seemed to have little impact on the male HA group's motivation. In contrast, student assessment had a very strong relationship with motivation. Motivation was related to positive peer-assessment ($r=.687$, $p<.05$) and self-assessment ($r=.765$, $p=.01$) in the pre-test, negative peer-assessment ($r=.636$, $p<.05$) and self-assessment ($r=.756$, $p<.05$) in the post-test.

Concerning the motivation factors, it was surprising that FA did not have any impact on *Learner*. *Language* was more strongly related to FA, followed by *Learning Environment*. *Language* was strongly related to open-ended questioning ($r=.814$, $p<.01$), positive peer-assessment ($r=.800$, $p<.01$) and self-assessment ($r=.672$, $p<.05$) in the pre-test, and suggestive feedback ($r=.648$, $p<.05$) and positive peer-assessment ($r=.719$, $p<.05$) in the post-test. *Learning Environment* was strongly related to self-assessment ($r=.777$, $p<.01$) in the pre-test, sharing learning aims ($r=.824$, $p<.01$), negative peer-assessment ($r=.683$, $p<.05$) and self-assessment ($r=.711$, $p<.05$) in the post-test.

HA boys were motivated by self-assessment in both the pre-test and the post-test. Peer-assessment, whether positive or negative, played an important role in their motivation too. Student assessment was a more important type of FA strategy than teacher assessment for HA boys. In terms of motivation factors, *Language* and *Learning Environment* had fewer but stronger correlations with FA in HA boys than LA boys. FA did not have any impact on *Learner* in HA boys.

Girls' LA group

The LA girls' group showed the most powerful relationship between FA and motivation among all the sub-groups. There were four FA strategies that had significant and strong correlations with motivation in the pre-test and the post-test. These were sharing learning aims ($r=.928^{\text{pre-test}}/.745^{\text{post-test}}$), open-ended questioning ($r=.664^{\text{pre-test}}/.744^{\text{post-test}}$), suggestive feedback ($r=.727^{\text{pre-test}}/.677^{\text{post-test}}$), and positive peer-assessment ($r=.724^{\text{pre-test}}/.738^{\text{post-test}}$). Besides, corrective feedback was strongly related to motivation in the pre-test ($r=.628$, $p<.05$), and close-ended questioning was strongly related to motivation in the post-test ($r=.843$, $p=.001$). It was teacher assessment and positive peer-assessment that had bigger

impacts on female students in the LA group than self-assessment did, which had no impact on motivation in general.

With regard to the motivation factors, there was not much difference between *Language* and *Learner*. Both of these factors were strongly correlated with sharing learning aims, questioning, suggestive feedback and positive peer-assessment. The strength of the relationship was slightly weaker in the post-test, but still remained strong. In the pre-test, there was another FA strategy that was strongly related to these two factors: self-assessment. In the post-test, the strong relationship disappeared for *Language* ($r=.685$, $p<.05$) and *Learner* ($r=.618$, $p<.05$) but emerged for *Learning Environment* ($r=.635$, $p<.05$).

Girls who were lower achievers were motivated by varied FA strategies, mainly in teacher assessment and positive peer-assessment. *Language* and *Learner* were two equally important motivation factors to them. Unlike the other sub-groups where self-assessment played important roles, self-assessment had no impact on female LA students' motivation.

Girls' HA group

For the HA girls, there were fewer correlations than in the LA girls' group. Although the number of the correlations was fewer, they were strong. The FA strategies that had correlations with motivation changed from close-ended questioning ($r=.686$, $p<.01$) and confirmative feedback ($r=.539$, $p<.05$) in the pre-test to corrective feedback ($r=.499$, $p<.05$) and positive peer-assessment ($r=.846$, $p<.001$) in the post-test. Among other strategies, self-assessment was the only one that had a strong correlation with motivation in the pre-test ($r=.635$, $p<.01$) and in the post-test ($r=.851$, $p<.001$).

In terms of motivation factors, *Language* and *Learner* were closely related to FA in the pre-test, but yielded fewer relationships in the post-test. In the pre-test, close-ended questioning, positive peer-assessment and self-assessment were the FA strategies with the strongest relations with *Language* and *Learner.* However, close-ended questioning only had a moderate relationship with *Language* in the post-test. Positive peer-assessment and self-assessment not only maintained the significant correlations with *Language* and *Learner*, they also generated new strong correlations with *Learning Environment*.

The girls' HA group had fewer correlations between FA and motivation than the LA girls' group, but these correlations were strong. More FA strategies could motivate LA girls initially, but later only confirmative feedback, positive peer-assessment and self-assessment could. Compared with other FA strategies, positive peer-assessment and self-assessment could motivate LA girls better, especially in the post-test. LA girls' motivation was influenced by these two FA strategies the most. Very similar to the picture for many other sub-groups, *Language* and *Learner* were significantly influenced by FA strategies, but only in the pre-test.

The correlations between FA and motivation for different genders and achievement levels are summarised below in terms of correlation features, pre-test/post-test, FA strategies and motivation factors:

1) Most of the correlations were moderate in the MA group, while the other two achievement groups had significant and strong correlations between FA and

motivation. The correlations indicated that the impact of FA on motivation was bigger in LA and HA. Among all the sub-groups, the girls' LA group had the strongest correlations between FA and motivation; the boys' LA group had the largest number of correlations between FA and motivation, but only half of these relationships were strong; meanwhile, the boys' HA group had the weakest correlations. The number and the strength of significant correlations in girls were greater than in boys. Girls' motivation was more readily influenced by FA than boys' motivation.

2) Most of the sub-groups did not show significant changes between the pre-test and the post-test; only the MA group had stronger or more correlations in the post-test, regardless of gender. However, most of the relationships in MA were moderate.
3) There was one FA strategy that did not have any relationship with motivation in all the girl groups, but was significantly correlated with motivation in all the boy groups, and that was negative peer-assessment. Therefore, this strategy was not recommended for girls, but it might be considered useful for boys. The most effective FA strategy for everyone was self-assessment, except for LA girls who were not affected by self-assessment at all. Other than this, positive peer-assessment was highly effective and well-maintained during the fieldwork in most of the sub-groups excepting HA boys. With this information, different strategies could be planned and applied to different sub-groups in a pertinent way.
4) *Language*-related motivation factors were the most important for students' learning. The *Learner* and *Learning Environment* factors showed different trends between the pre-test and the post-test. The importance of *Learner* was slightly reduced after a period of study. *Learning Environment* showed an insignificant increasing impact during the fieldwork.

4.6.6 Summary of relationship between FA and motivation

Table 4.7 summarises the significant impact of FA on motivation as well as each of the three motivation factors in LA and HA. Because the impact of FA on MA's motivation was moderate or weak, it is not listed in the table. As shown, FA strategies are the most useful for LA girls to boost their motivation. On the other hand, FA had the smallest impact on motivation in HA boys. However, for these boys, *Learning Environment*-related motivation strategies engaged them relatively well and this might be a good angle to improve their language learning.

Table 4.7 Summary of the impact of FA on motivation

Impact between	Size of the impact in decreasing order			
	1	2	3	4
FA and motivation	LA girls	LA boys	HA girls	HA boys
FA and language	LA girls	HA girls	LA boys	HA boys
FA and learner	LA girls	LA boys	HA girls	HA boys
FA and learning environment	HA boys	HA girls	LA boys	LA girls

4.7 Relationship between FA and achievement

To answer the research question '*Is there a relationship between FA and students' achievement?*', the correlations between FA and achievement were calculated in SPSS.

4.7.1 In general

Table 4.8 sets out the Pearson correlation values between FA and achievement and shows that most of the values were under .300, indicating relationships that were significant but small. The use of negative peer-assessment had no significant relationship with achievement.

Table 4.8 Correlation between use of FA and achievement

		Average achievement
Use of Sharing Learning Aims	Pearson Correlation	.172*
	Sig. (2-tailed)	.016
Use of Open-ended Questions	Pearson Correlation	.290**
	Sig. (2-tailed)	.000
Use of Close-ended Questions	Pearson Correlation	.276**
	Sig. (2-tailed)	.000
Use of Confirmative Feedback	Pearson Correlation	.270**
	Sig. (2-tailed)	.000
Use of Suggestive Feedback	Pearson Correlation	.205*
	Sig. (2-tailed)	.004
Use of Corrective Feedback	Pearson Correlation	.189*
	Sig. (2-tailed)	.008
Use of Positive Peer-assessment	Pearson Correlation	.246**
	Sig. (2-tailed)	.001
Use of Negative Peer-assessment	Pearson Correlation	.046
	Sig. (2-tailed)	.525
Use of Self-assessment	Pearson Correlation	.256**
	Sig. (2-tailed)	.000

** Correlation is significant at the .001 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

4.7.2 Achievement group

If grouped by achievement level (Appendix, Table 8), only three FA strategies had significant correlations with LA and MA, and these were: sharing learning aims with LA (r=.344, p<.05), corrective feedback with LA (r=.415, p<.05) and positive peer-assessment with MA (r=.240, p<.01).

4.7.3 FA environment

If grouped by FA environment (Appendix, Table 9), similar to the case in the previous section, there was no strong correlation between FA and achievement. In FAR students, the most effective FA strategy for improving achievement was confirmative feedback (r=.402, p<.001); while in FAL students, it was open-ended questioning (r=.312, p<.01). Sharing learning aims

and negative peer-assessment did not show strong correlations with achievement in FAR nor FAL.

4.7.4 Gender

If grouped by gender (Appendix, Table 10), similar to the case in the last two sections, there was no strong correlation between FA and achievement. For boys, the strongest correlation was between close-ended questioning and achievement ($r=.381$, $p<.001$). For girls, the most useful FA strategy to improve achievement was confirmative feedback ($r=.383$, $p<.001$). Although weak, the correlations between FA and achievement were greater in boys than in girls.

4.7.5 Matrix comparison of Achievement group and Gender

If grouped by achievement and gender (Appendix, Table 11), unlike other comparisons, LA girls showed significant and strong correlations between FA and achievement, especially in teacher assessments. Sharing learning aims ($r=.692$, $p<.01$), open-ended questioning ($r=.763$, $p<.01$), confirmative feedback ($r=.608$, $p<.05$), suggestive feedback ($r=.641$, $p<.05$), and corrective feedback ($r=.594$, $p<.05$) were all significantly related to LA girls' achievement. Besides, close-ended questioning had a strong relation with HA boys' achievement ($r=.604$, $p<.05$); positive peer-assessment had a weak relation with MA girls' achievement ($r=.267$, $p<.05$). Therefore, FA had big impacts on LA girls' achievement, but not on other sub-groups' achievement.

4.7.6 Summary of relationship between FA and achievement

Considering all the tests above, the impact of FA on achievement was significant as well as strong in LA girls only. Although there was no particularly strong correlation between FA and achievement in other groups, the correlations were statistically significant.

4.8 Relationship between motivation and achievement

To answer the research question '*Is there a relationship between students' motivation and achievement?*', the correlations between motivation and achievement were calculated in SPSS.

4.8.1 In general

Table 4.9 shows the correlation between different motivation factors and achievement. In general, motivation and achievement had moderate correlation in the pre-test ($r=.387$, $p<.001$) and weak correlation in the post-test ($r=.236$, $p<.01$). The motivation factor *Learning Environment* did not have a significant relationship with achievement. The strength of correlation between *Language* and *Learner* factors and achievement was moderate ($r=.472/.499$, $p<.001$) in the pre-test, but weakened in the post-test ($r=.373/.386$, $p<.001$).

Therefore, motivation was significantly correlated with achievement, but the strength of the relationship was just moderate. The stronger the motivation was, the better the achievement would be. Over the course of the study, the impact of motivation factors on achievement became smaller and smaller. The most significant element in motivation was *Learner*. *Learner*-related motivation played a bigger role in students' achievement than other factors, while *Learning Environment* showed no significant relationship with achievement.

Table 4.9 Correlation between motivation and achievement

		Pre-test	Post-test
Language	Pearson Correlation	.472**	.373**
	Sig. (2-tailed)	.000	.000
Learner	Pearson Correlation	.499**	.386**
	Sig. (2-tailed)	.000	.000
Learning environment	Pearson Correlation	.020	-.068
	Sig. (2-tailed)	.784	.344
Motivation	Pearson Correlation	.387**	.236**
	Sig. (2-tailed)	.000	.001

** Correlation is significant at the .001 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

4.8.2 Achievement group

If grouped by achievement (Appendix, Table 12), motivation and achievement were not related in HA students. The MA group showed small correlations between motivation and achievement ($r=.228^{\text{pre-test}}/.215^{\text{post-test}}$), as well as *Learner* and achievement ($r=.276^{\text{pre-test}}/.287^{\text{post-test}}$), and *Language* and achievement ($r=.198^{\text{pre-test}}/.207^{\text{post-test}}$). For LA students, motivation had a significant and strong correlation with achievement in the pre-test (r=.520, p=.001). All the motivation factors also had moderate relationships with LA's achievement in the pre-test. However, in the post-test, there was no significant correlation.

The LA group's motivation and achievement showed mutual influence in terms of every motivation factor at the beginning of study, but the effect would disappear as time passed. The MA group's motivation and *Learner* were related to achievement; the impact was increasing over time. For HA students, their motivation and achievement were not related.

4.8.3 FA environment

If grouped by FA environment (Appendix, Table 13), FAL showed stronger correlations between motivation and achievement. There were only two significant and strong correlations: *Language* (r=.536, p<.001) and *Learner* (r=.566, p<.001) with FAL's achievement. The strength of other relationships was either moderate or small. All the correlations were weakened in the post-test except for FAR's achievement and *Learner* ($r=.374^{\text{pre-test}}/.400^{\text{post-test}}$). FAL's motivation and achievement correlations had a greater decline than FAR's.

In terms of motivation factors, *Learning Environment* had no relationship with achievement. The correlations between achievement and *Learner* were slightly stronger than between achievement and *Language*.

Motivation had a bigger impact on achievement in FAL, especially in the beginning of study. Motivation and achievement were correlated in FAR, but the impact was not as great as in FAL. FAL students may perceive achievement as more important than FAR students do. They may have enjoyed the feeling of achievement or have higher standards for themselves before they start to study; that could explain why their motivation derived from the *Learner* factor was significantly correlated with achievement in the pre-test. To improve achievement in FAL, promoting students' motivation would be an effective method, *Learner*- and *Language*-related motivation in particular.

4.8.4 Gender

If grouped by gender (Appendix, Table 14), the data showed that boys had stronger correlations in the pre-test (r=.463, p<.001), but girls had stronger ones in the post-test (r=.326, p=.001) due to there being no significant correlation between motivation and achievement in boys in the post-test. Boys' motivation-achievement correlations were weakened during the fieldwork; conversely, girls' were stronger. This indicated that boys' achievement did not depend on their motivation along the time of the study, while girls' achievement grew more dependent on their motivation. In all the significant relationships, only one was strong, which was between *Learner* and boys' achievement (r=.584, p<.001) in the pre-test. Therefore, the impact of motivation on achievement was significant, but not strong.

In terms of motivation factors, as in the previous section, *Learning Environment* had no relationship with achievement. There was no significant difference in the boys' and girls' achievement-*Language*/*Learner* relationships.

The impact of motivation on achievement was sizeable for boys in the pre-test. Over the period of the study, boys soon lost their interest in achievement and their engagement by motivation strategies; meanwhile, girls continued to pay more attention to achievement. Therefore, to help boost achievement in boys, it would be important to sustain their interest in the connection between motivation and achievement.

4.8.5 Matrix comparison of Achievement group and Gender

If grouped by achievement and gender (Appendix, Table 15), it was clear that among all the sub-groups LA girls' achievement and motivation were the most significantly correlated. Motivation was strongly related to their achievement in the pre-test (r=.712, p<.01), but there was no correlation in the post-test. In the pre-test, LA girls' achievement was strongly related to *Language* (r=.700, p<.01) and *Learner* (r=.755, p<.01), but in the post-test, their achievement was strongly related to *Language Environment* (r=.605, p<.05) only. There were a few moderate or weak relationships between MA boys' achievement and motivation, such as motivation in the pre-test (r=.269, p<.05), and *Learner* ($r=.344^{\text{pre-test}}/.253^{\text{post-test}}$).

Learner also had a moderate correlation with MA girls' achievement in the post-test (r=.340, p<.01).

It was obvious that the impact of motivation was great on LA girls' achievement. However, the impact seemed to derive from *Language* or *Learner* aspects in the beginning, then changed to *Learning Environment* later. In the post-test, the overall impact of motivation on achievement disappeared. Besides these patterns, the MA boys' achievement was easily affected by *Learner* aspects too.

4.8.6 Summary of relationship between FA and motivation

The findings above showed that there were some relationships between motivation and achievement, but the correlations were not striking. Some groups showed relatively strong correlations, such as LA, FAL and girls. The impact of motivation on achievement is apparently slightly greater in these groups. It is worth highlighting that LA girls' achievement was the most readily affected by motivation.

4.9 Summary of quantitative findings

In this chapter, I discussed the findings from the quantitative data obtained from the questionnaire about motivation, the survey about FA and the achievement scores collected from quizzes. After introducing the basic background information about the data, I analysed each theme in this study: achievement, motivation and FA. The subsequent sections discussed the relationship between each two themes to answer the research questions: '*Is there a relationship between FA and motivation, between FA and achievement, and between motivation and achievement?*'.

It emerged that there were significant correlations between each two themes, but the strength was moderate. In different group contexts, there were strong correlations between sub-groups. For instance: the interrelationship between FA, motivation and achievement was great in girls with lower achievement; the relationships between achievement and the motivation factors *Language* and *Learner* were stronger than with *Learning Environment*; positive peer-assessment and self-assessment were the most effective strategies that significantly correlated with motivation and achievement. It was therefore valuable to dig into the results and analysis to identify the FA strategies best suited to promote a particular aspect of motivation or to boost achievement.

In the following chapter, I will analyse the qualitative data on FA and motivation from interviews. This analysis will add to and further explain the picture drawn by the quantitative data.

Chapter 5 Qualitative Data Analysis

5.1 Introduction

The qualitative data for the main study were collected from semi-structured interviews with two teachers and twelve students. The two teachers came with different assessment habits and the twelve students were from different achievement levels. In the following section, the findings and the analysis are illustrated in the order of FA strategies. The information was collected in Chinese. To interpret meanings and subtexts with accuracy to the largest extent, the data were firstly coded and analysed in Chinese with NVivo 11.3, then translated into English for the quotations. A sample of the Chinese transcripts (Appendix 12) and its analysed English translation (Appendix 12.1) with codes are attached.

The interviews focused on teachers' and students' perceptions of FA. These would address the research question 4: What is the impact of **FA** from English **students'** perspective? (5.2) and research question 5: What is the impact of **FA** from English **teachers'** perspective? (5.3).

In the following section, the student interviewees will be referred to as:

Gender	Boy	Girl	Boy	Girl
FA / Achievement	FA-rich environment		FA-lacking environment	
Higher achiever	HBR	HGR	HBL	HGL
Medium achiever	MBR	MGR	MBL	MGL
Lower achiever	LBR	LGR	LBL	LGL

The teacher interviewees will be referred as RT (teacher who created FA-rich environment) and LT (teacher who was in FA-lacking environment).

5.2 Students' perceptions of FA

5.2.1 Sharing learning aims

The teachers used the strategy sharing learning aims (SLA) either at the beginning or the end of a class. RT would use PowerPoint but LT preferred handwriting on the blackboard. Five interviewees from FAR and one from FAL mentioned that the teacher used SLA often; it seemed RT used it more often than LT.

It is gratifying that all students found the aims clear, but they saw the level of difficulty differently. Most of the HA or MA students found the aims easy; meanwhile, all LA students admitted that the aims were "*sometimes easy, sometimes hard*". Two MAs from FAL said that they may feel "*burdened*" or "*fed up*" with the difficult aims.

There were two types of learning aims, 'what-to-learn' and 'what-to-do'. The different types of lessons had different aims - that the new lesson would talk about "what-to-learn" and the exercise/practice lesson would talk about 'what-to-do' since students would already know

the content. Students who recalled 'what-to-learn' aims during the interview all liked SLA, but those who recalled 'what-to-do' aims said that it was the content of the aims that determined whether they liked SLA or not. They welcomed the aims with language learning, but disliked the aims involving tasks, such as dictation or correcting the mistakes in homework. Because of this dislike, they queried the need for such aims for each lesson. In light of this attitude towards aims with tasks, I would like to suggest that the aims with language learning are preferred.

Most students held positive attitudes towards the SLA strategy and the reasons were various. Only HBL strongly disagreed with this. He was a highly self-motivated student who argued that:

> *HBL: [one] should have a warm heart to English, see it as a part of your life, you are willing to learn English.*

For others, knowing the aims could help them know "*what to learn*" "*in advance*", and decide whether the lesson was "*difficult or easy*". Knowing the aims, they felt "*curious*" about what was to be learned next and found it was "*easier to learn*" with "*clear targets*".

> *MGR: Because this will mobilise our initiative, let us take the initiative to learn English, let us know what the aims are today, we will have the motivation to learn.*

The commonest reason for finding the aims helpful was the content of the aims. Students wanted to know whether the new knowledge would be interesting or boring, complex or simple, difficult or easy. The contents of the aims determined their willingness and enthusiasm to continue.

> *LGL: Because there is new knowledge, we are curious, we can listen to it and see what is new, we learn at the same time.*
>
> *LGR: Because we learn learning aims through teachers, we can know what we are learning about in this lesson. But also know that this lesson is difficult to learn or not … Because, em, what the teacher said, can let us know whether this lesson, this lesson is interesting or not.*

The sense of direction was also important. Nearly half of the interviewees admitted that knowing the learning aims could tell them "*where to go*" and they would not feel "*blind*" or "*lost*".

> *HBL: Like it, because there is an aim, otherwise, we [feel] confused and do not know what to learn.*
>
> *HGL: Because sometimes you learn blindly, but you do not know what to learn. And then the teacher gives you a designated aim, you can follow this goal and go for it.*

However, MA students expressed their worries about the aims. They listed some reasons why they may not like it: "*too difficult*", "*too hurried*" and "*too much information*".

> *MGL: Want to learn English more, that is, under the circumstance that we can understand [the aims].*

They clearly knew what they do not want. This can help the teachers to avoid the traps in designing learning aims. Students tend to like learning aims which are suitable to their achievement levels. When designing aims, teachers would be better to set one aim but at

different levels. Therefore, students could do their best to achieve the highest level on the aim 'ladder' according to their own situation. The teachers also need to take enough time to explain the aims clearly and make sure that they are written in a simple and straightforward way.

All of the interviewees believed that they could improve their achievement via SLA. Unfortunately, boys from MA and LA all said that they did not have a reason to support their opinion. For other students, the aims were the highlights of their study, and their reasons have already been summarised, *e.g.* easy to "*make notes*" and "*recite*". The aims pointed out the direction of study, and could guide or "*lead*" the students to achieve more.

> *HBR: If you have more aims, then you will work very hard to learn, if you have fewer learning aims, [students] will take it easy… [if you] work harder, the more you will achieve.*

To sum up, sharing learning aims was definitely an effective FA strategy in students' perspective. It could motivate students and help them with achievement because it provides them with a sense of direction which they can work hard towards. They knew what to learn and were aware of the level of difficulty in the upcoming lesson. Meanwhile, having a clear picture in mind in advance of learning the content may stimulate curiosity about the content.

5.2.2 Questioning

Teachers used questioning differently. Five out of six interviewees from FAR confirmed that their teacher used open-ended questioning more often than closed questions, and four out of six from FAL said that their teacher used more close-ended questioning than open-ended questions. HGL and MGL thought that LT used both questioning methods equally. In most of the cases that teachers used open-ended questioning, it was at the beginning of a lesson, which normally would be for the daily greetings, a review of the last lesson and an introduction to the new lesson. HBR mentioned another case:

> *HBR: For example, once we finish reading the story, Miss will ask us about the people, who are they, where are they and what did they do.*

It was clear to deduce from students' perspectives that RT used open-ended questioning more often, while LT preferred close-ended questioning. Both teachers explained their perceptions and these are reported in 5.3.2.

Three quarters of the interviewees preferred open-ended questioning, because it could help a student "*improve English*" and "*open the mind*".

> *HGR: Because it can open my mind.*
>
> *HGL: to improve your ability to communicate in English.*
>
> *MBR: just, others' answer may open your mind, then you can continue.*
>
> *LBL: with long sentences, you will know more.*

These challenging questions offered "*more possibilities*" in the answer, which allowed students "*express freely*", "*say whatever you want*", because they had no "*fixed answer*".

One third of the interviewees preferred close-ended questioning. Their only reason was that it was "*easy to understand*" and "*easy to answer*". It would not take them too long to answer a close-ended question.

In all the responses, HBR was a very interesting case. He insisted at first that close-ended questions were more difficult:

> *R (reporter): Why [do you prefer] close-ended questioning?*
> *HBR: Because it is more difficult… The questions are comparatively difficult.*
> *R: Could you give me an example of close-ended question?*
> *HBR: [thinking for 10 seconds, then waved the head]*
> *R: I have an example for you, do you like this camera? Camera means zhaoxiangji (camera in Chinese). Do you like this camera? How will you answer it?*
> *HBR: Yes.*
> *R: This is a type of close-ended questions.*
> *HBR: So it is more difficult.*
> *R: Okay. It is more difficult then.*
> *HBR: The word camera is difficult.*
> *R: What if it is "do you like this pen"?*
> *HBR: It is easy once you learn it.*

It seemed that HBR thought close-ended questions were difficult only because of the new word. However, there was another conversation about the impact of questioning on achievement:

> *R: Which one do you prefer if you know all the words?*
> *HBR: Open-ended questions*
> *R: Open-ended questions, um? Why?*
> *HBR: [You will] speak more in open-ended questions, it can improve our English.*
> *…*
> *R: So, do you think, the two ways of questioning, which one can improve your English more, or improve your ability in English learning?*
> *HBR: Close-ended.*
> *R: Still close-ended, okay, why?*
> *HBR: It must be more difficult with less language.*

He had a strong and positive attitude to close-ended questioning but he could not fully explain this. I suspected he mixed up the two ways of questioning because normally students see open-ended questioning as more difficult. However, after giving the example, he still insisted on his opinion. He said he preferred open-ended questions because it can help improve English; yet, this way of questioning does not motivate him. He also said close-ended questioning can improve his English more, even though he had just claimed that he preferred open-ended questioning for the same reason. Interviewees in one-to-one interviews are rarely challenged by the interviewer - they might say something inconsistent with earlier responses or that patently could not be true (Bryman, 2016). In this case, HBR's answers were inconsistent, which made the data difficult to analyse.

HBL, the highly self-motivated student, also thought that questioning would not motivate him. His answer was similar to the one in 5.2.1,

HBL: I think, [students] themselves should love English. I think [both of the ways of questioning] are not helpful.

He had strong motivation in language learning, therefore it was understandable that from his viewpoint FA strategies were less important.

There were nine interviewees who believed that open-ended questioning was more helpful in improving achievement. The main reason was that varied answers to questions can bring up more ideas and increase the chances of being challenged and tested.

HGR: the teacher asked you an open-ended question, there are many answers, which means everyone has the chance to answer it. Then you will answer it in your own way, then, the teacher will, sometimes, write others' answers on the board, you can choose any of the answers. Keep learning and it will improve your achievement, you will know more.

MBR: In open-ended questions, this [answer] may not be tested in the exams, but others' [answers] could be.

On the other hand, according to HBL and HGR, close-ended questions may be "*too simple*" for them. Students in HA enjoyed the greater challenge.

In summary, most of the students found questioning important in both motivation and achievement, open-ended questioning in particular. Open-ended questioning could improve their achievement due to variety in the answers. It also opened their minds when students could express their own opinions freely.

5.2.3 Feedback

According to the literature review, the function of feedback was categorised into three general types: confirmative feedback, suggestive feedback and corrective feedback – all forms of verbal feedback. In fact, non-verbal feedback also played an important role in the class according to the interviews, for FAR in particular. HBR, MGR and LBR gave some examples of the combinations of different feedback:

HBR: A few body language feedbacks and normally the teacher uses verbal feedback, like good... If the teacher is about to criticise the students, she will use body language, such as turn stern-faced, with an unhappy face, then let the student sit down and ask another to answer the question. [If the teacher is about to compliment the students,] she will say the words with positive meanings, such as good.

MGR: If you were correct, Miss will use some terms, like very good, or some encouragements, also give the thumb up, or have a hug. If some students cannot answer, Miss will say, it is okay, sit down first, think about it again, I will ask you later.

LBR: If they answered well, hugs, if not, then [the teacher] will encourage them. [The teacher does not use corrective feedback, instead she] will ask others to answer the questions.

According to the interviews quoted above, RT used different forms of feedback, for example, body language, facial expression, re-framing an answer, suggestions from peers, simple confirmation, encouragement, comforting, correction.

RT showed that she paid attention to students' emotions. No matter how the students' performance was, RT was good at making them comfortable and inspired by different kinds of feedback. On the other hand, LT used fewer forms of feedback:

> *HGL: Good, wonderful, then sit down, nothing else.*
>
> *MBL: Good/dui (Correct in Chinese)/very done*.*
>
> *MGL: Good if you are good; or a sticker on the achievement sheet** if you are really good… If you behaved badly for three times, a sticker will be taken off… [R: What will the Miss say after your answer?] just no comment.*
>
> *LBL: If good, thumbs up, nothing else.*
>
> *LGL: [Miss] will say very good…Zai jie zai li (Keep on working in Chinese) if someone is not good.*
>
> *I think he meant "well done".
>
> **This was a credit system presenting everyone's performance with stickers. It was on the back wall of the classroom.

LT mainly used simple verbal confirmation, such as "*good*, *very good*, or *wonderful"*. Students also mentioned that she used a credit system, rhetorical questioning, *etc*. There was a big difference between the two teachers in terms of assessment habits. As reported in 5.3.3, the teachers talked about their experience and thoughts in using feedback.

When talking about feedback and motivation, interviewees who were motivated by confirmative feedback listed the following reasons: "*feel happy and encourage*d", have "*more interests*" to learn and "*more confidence*" because of "*the recognition of my efforts*". Suggestive feedback was popular too because it could help students to "*think on their own*". No one thought corrective feedback would motivate them.

> *HBL: Corrective feedback is definitely inappropriate, because the teacher is telling the answer directly, just like doing the homework for you. You do not know how to answer, the process, the reason, why you think in this way. I prefer the teacher telling me the way of thinking if you have a problem.*
>
> *HGL: If it is the corrective feedback, they [the students] will rely on their [the teachers'] answers, rather than themselves.*
>
> *MBL: If correct us directly, we will listen to the answer directly, we would not think on our own later.*
>
> *MGL: Also I do not like the type which tells you the answer directly. From my heart, it does not feel like my own answer… It feels that [the teacher] have given up on you".*

Four interviewees from FAL pointed out that corrective feedback cannot motivate them, for the reasons that it was easy to "*rely on others*" and a student "*cannot think on one's own*". MGL even felt abandoned if the teacher used corrective feedback on her.

Almost all FAL interviewees believed that suggestive feedback could help them the most in their achievement. It was easier to "*remember how you learned it*" via "*thinking on your own*". Half of the interviewees from FAR agreed that confirmative feedback brought the strongest motivation to their study because it was inspiring.

HBL: Because you need to use the suggestions to think on your own. This also could improve your ability of thinking.

MGR had a unique but valuable viewpoint. She pointed out that confirmative feedback cannot be 'too much'.

MGR: [the people who receive too much confirmative feedback] will have a sense of conceitedness, in that case, he will not work hard on English anymore… Too much confirmative feedback may make one too big, too strong and proud. Anyway, [one] has very high self assessment. Then, s/he would not motivate himself/herself to learn more. If so, s/he gradually feels that s/he is excellent in English, there no need to learn more. Therefore, I think too much [positive] feedback is not that good.

As described in 4.6.1 and 4.6.4, analysis of quantitative data revealed that confirmative feedback had the weakest relationship with motivation, especially in the girls' group. MGR's viewpoint may explain the reason why confirmative feedback was not important in motivating students. Students may enjoy confirmative feedback at first, but due to reasons such as the amount or form of the praise, they gradually take confirmation for granted. They still enjoy the praise, but it would not motivate them as much as before. Students explained that although they still enjoy the praise, it may have the effect of changing students' motives from an interest in language and confidence in themselves to a desire to earn recognition from teachers. This is a reminder to teachers to be careful not to over-use any strategy which confirms, affirms or praises a student's performance, even if the students like it.

To summarise, confirmative feedback and suggestive feedback were much more popular than corrective feedback. Students highlighted the confidence and encouragement gained from confirmative feedback, as well as learner autonomy and inspiration from suggestive feedback. In contrast, corrective feedback, which was used frequently in the class, produced negative responses. There was another very interesting finding that non-verbal feedback was more important than expected. It emerged that seemingly minor aspects of teachers' behaviours in class, such as facial expressions or tone of voice revealing certain attitudes, had a big impact on students.

5.2.4 Peer-assessment

The FA strategies discussed above were mainly initiated by teachers. By way of contrast, peer-assessment and self-assessment require students to participate more in the strategy. Three quarters of the interviewees said that they did not receive or give peer-assessment and half of them said they did not use self-assessment. This was also evidenced in Figure 4.11. Peer-assessment and self-assessment were used much less than other strategies. Some interviewees even claimed that they had never seen this type of assessment before.

According to the traditional teaching and learning style and the Confucian culture in China, teachers tended to dominate the classroom. Even though educators in China are trying to change this situation, it seems the influence of the old style persists. Students were not accustomed to FA strategies largely led by themselves. Besides this, they may have been too young to notice that the activity they performed was actually peer-assessment. However,

they were looking forward to having more peer-assessment, particularly suggestions from peers:

> *HBL: For example, you [peers] can give me some comments after I read the passage, ah, 'you need to improve at blah blah blah', 'you are good at blah blah blah' and 'you need to make it louder or quieter', something like facial expression or gestures.*

Sharing experience was another type of peer-assessment they looked forward to:

> *MGR: If we did well, then we will have a talk, sharing the ideas, something like how to improve reading.*

Applause was a form of peer encouragement experienced by students in FAR, although it was non-verbal and was often led by teachers. Students said that they "*act mechanically when applauding for others*". However, the students who received the encouragement did not know that, so it would not affect how they feel about the compliment. It is normal that people feel encouraged and motivated by compliments. Therefore, even if the positive peer-assessment was just a classroom routine, its impact cannot be dismissed.

Three quarters of the interviewees agreed that peer-assessment could motivate them in language learning, for the reason that "*it is hard to notice your own faults*", or if the peer-assessment is "*simple*". However, some students disagreed because "*everyone's standards are different*" and "*it will feel awkward*".

They gave further detailed reasons about why they were motivated by each type of peer-assessment. It seemed that negative peer-assessment was more popular than positive peer-assessment. All LA interviewees and three MA and two HA students believed that negative peer-assessment was helpful. LA interviewees were willing to "*work harder*" or "*correct the mistakes*" to "*prove themselves to others*". HA and MA interviewees were inclined to "*reflect*" on themselves with "*new aims*" and more suggestions.

> *HBL: Because it [negative peer-assessment] will give me something to chase, louder, more vivid, or more fluent, there is always a new aim.*
>
> *MGR: …why I cannot answer the question, it must be me playing around unconsciously, or, or it is due to one of my faults.*

Meanwhile, HA and MA also worried about the unpleasant effect peer-assessment may bring to students. For instance, negative peer-assessment may "*hurt one's pride*" and make the student "*feel depressed*". As HGL and MGL said in their interviews:

> *HGL: Because I think the negative ones, for the most of time, pupils are too straightforward, then what they said may hurt one's pride… whereas the negative peer-assessment sometimes may let you feel depressed.*
>
> *MGL: It is just sometimes when you read, when [others] claim that I am not good, you know that, like others say bad things about you, you will sad too… like I said before, [self-assessment] brings me confidence. I do not need others to judge me. Peer-assessment will bring me burdens.*

Consequently, it was HA and MA who preferred positive peer-assessment. They described it as "*inspiring*" and encouraging. Positive peer-assessment could point out where they did

well and remind them to sustain their gains. HBL also warned people of complacency in case "*everyone is saying you are good at everything*".

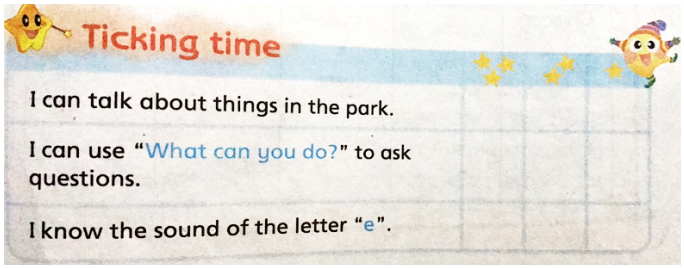
Compared to HA and MA, interviewees in the lower achievement group were more eager to prove themselves to others. Possibly because their confidence may not be high due to their achievement level, they reacted more strongly to negative peer-assessment. It seems they need closer attention and more careful treatment by their peers (and teacher).

Of course, positive peer-assessment cannot solve every problem. LA students were more easily motivated by negative peer-assessment, while HA and MA preferred positive peer-assessment. When it came to achievement, the impact of both positive and negative peer-assessment was similar. Half of the interviewees believed that positive peer-assessment could improve achievement, while the other half said the opposite. Positive peer-assessment may encourage students and spur them to correct their mistakes or keep up their advantage; on the other hand, negative peer-assessment may challenge their confidence then make them force themselves to improve, as well as "*leave a deeper impression of the knowledge*".

Peer-assessment was a useful assessment method but it was not being adopted fully. With the development of student-oriented learning, teachers might think they will lose some of their control in the classroom. The influence from peers will become stronger and stronger. Positive peer-assessment had great impact on motivation and achievement, especially in HA and MA groups. LA students welcomed negative peer-assessment, but may not notice that they are hurt or negatively affected by this input. Compared to students at more advanced levels, such as in higher education or senior school, students in primary school are comparatively naïve and fragile. They may suffer more from peers due to their immature egos, but conversely the positive impact of peers' assessment may be magnified. Therefore, more training or guidance on effective peer-assessment is recommended for teachers and students to make better use of it.

5.2.5 Self-assessment

As mentioned above, self-assessment was not often used in the classes studied. Six interviewees, *i.e.* half of them, admitted that they did not use it, and five of these were boys. Those students who did use it reported that the types of self-assessment most often used were encouragement and reflection. It was mainly the girls who were using them. MGR commented that a good way to reflect on oneself was "*to set a learning aim for today, if [I] did not reach the target, think about where I was short at*". This method was similar to the textbook self-assessment forms that HBL mentioned. In the textbooks they are using, there is a short self-assessment form called 'ticking time' which is appended to each unit (see an example in the picture below). This was a promising sign that showed that the local education authorities have already started to put emphasis on forms of assessment other than the traditional ones such as tests or teachers' feedback only.



Other FA strategies also had great influence on students' self-assessment. As HGL and MGL said:

> *HGL: Give myself a comment, do not feel too good about myself. See exactly how well I did, give myself a comment from deep inside. Compare to others' assessment, then improve.*
>
> *MGL: Sometimes, when the teacher confirms my answer, I will be confident in myself, then I will improve myself. If the teacher told me the answers straight away, I kind of dislike it. At that time, I will have very negative self-assessment.*

Accordingly, sharing learning aims, feedback, peer-assessment and self-assessment were closely correlated. Their impacts could affect each other. Therefore, to achieve a better combined, overall result , all the strategies need to be integrated with teaching and learning.

However, half of the interviewees disliked the self-assessment strategy or did not think it would motivate them in language learning, for reasons such as "*I do not know how to assess myself*" or "*people normally do not know their mistakes*". Those closely involved with work perhaps cannot see clearly, therefore, self-assessment activity needs more help from others, such as teachers' instruction, and a guide book about how to do self-assessment. The 'ticking time' activity mentioned above also needs more guidelines to help students gauge what is required to achieve one, two or three stars.

Students showed their belief in self-assessment from two angles, emotion and knowledge. For example, self-assessment could "*help me find the faults*" [knowledge], "*know myself better*" [emotion] and "*comfort myself*" [emotion] after receiving negative *peer*-assessment. With positive self-assessment, they felt "*proud*" [emotion], and with negative self-assessment, they would have "*deeper impression*" [knowledge] of the learning.

With all these positive responses, almost all interviewees strongly agreed that self-assessment was helpful for improving achievement. Only LGL was worried about it:

> *LGL: sometimes I read very well, but actually [I was] not, sometimes really bad, but it was good… [there is no] standard criteria.*

Other interviewees believed that self-assessment could improve the achievement if it was the "*appropriate amount*" and had clear aims. If faults were found, they will "*correct them, turn them into something good*". Because they found the faults or errors themselves, students showed a more positive attitude to correcting them than they did with peer assessment. It was easier for students to accept their own suggestions than others'. HBL even asked for more negative self-assessment.

HBL: I think I should give myself some negative assessment. That makes me better. There are always other able men behind an able man, [students] cannot think themselves are excellent, and then do nothing, [I] still need something like negative assessment.

LGR indicated positive psychological benefits in her self-assessment:

LGR: If I still need to work hard on this question, if it is tested in the exams, my scores will be higher.

R: What if it is not in the exam?

LGR: If it is not in the exam… But I have already known the knowledge, it is good too.

In general, self-assessment, a less often used FA strategy, was playing an important role in motivating students and improving their achievement. Interviewees enjoyed different forms of self-assessment. With the help of the local department of education and the efforts of educators, a basic self-assessment section was put into the textbook. However, the only concern raised related to adequate guidance, so it is recommended that the textbook includes more detailed instructions and regulations about how to use the self-assessment activities.

5.3 Teachers' perception of FA

The teachers in the current study came from the same school, but from two different school sites. RT, as indicated in the abbreviation, created an FA-rich environment for students; meanwhile, LT, for a range of reasons, used less FA in her class. Their personalities and experience led to different teaching styles; and their insights into education led to different interpretations of FA.

Teachers' personalities and teaching styles

LT mentioned in her interview that she had a relatively 'aloof' teaching style (in her own words) because she had a dilemma as a new teacher:

LT: I think as a new teacher, my experience is not enough. When I was learning psychology in education, it was said that the new teacher focuses on themselves at the first stage, then [they will have a stage of] focusing on the teaching situation, and the final stage is focusing on the students. I think in my first year, maybe I often care about myself… And then to the second year of work when I found myself paying more attention to the whole situation, that is, whether the whole teaching can be smooth or not.

These survival concerns, teaching situation concerns and pupil concerns were put forward as important issues in becoming a teacher by Fuller and Bown (1975). However, the fact that little FA was used in LT's class cannot be denied. Through lesson observation, I received the impression (and this appeared to be supported by some of the children) that LT smiled less often than RT, and generally her body language conveyed detachment. On the other hand, RT was demonstrably very warm hearted, always smiling, always hugging students. She mentioned "*love*" and "*understanding*" all the time in the interview.

RT: I think assessment was on the basis of love and understanding itself, so it will create a way of assessing behaviours… So the assessment itself, excepting the basis of love, it is also necessary in teaching… So I think the true assessment must be based on the love for children, it is just showing your viewpoint on a behaviour… I think if everyone can hold this attitude to students, that is why I can always receive love from my students.

In terms of interaction style, RT appeared to show a kind of maternal affection for her students. In RT's class, her voice was soft, and her tone was rather child-like; meanwhile, LT was more like an adult talking to other adults. Teachers' personalities could be one of the main reasons for the difference in teaching and interaction styles; life experience could be another.

Insights into education and perceptions of FA

LT's use of the strategies in class appeared to be quite rigid. She mentioned in the interview, with a humble attitude, that she would like to learn more about how to use each FA strategy in a better way. She also mentioned educational concepts such as *discovery learning* and *independent learner*. It is possible that LT's assessment strategies were fragmented due to her relative lack of experience. Without a holistic perspective, she did not bring all the strategies and follow-up activities together and understand them thoroughly. She said that she did not have time to dig deeper about why and when she might use these strategies. Judging from the literature in this area, she is not the only teacher to have shown lack of a clear understanding of the term 'formative assessment' (Xiao & Carless, 2013; Afitska, 2014b).

On the other hand, RT put FA into a broader picture. There were several key words in her interview responses: *love*, *independent*, *long-term* and *socialisation*. From her perspective, FA was not just a set of strategies to help students meet learning aims in her class, but a tool that she used to achieve goals which might extend beyond language learning. Her main concern was about how to nurture independent people via FA, so that the students could play an active part in the society, and adopt learning habits and values from the teacher and their classmates.

She thought that assessment was "*a recognition from teacher*", as well as "*a delivery of teacher's values*". She wanted to pass her values on to the students, especially her philosophy of life. In the following sections, teachers' perceptions in connection with the five FA strategies are described and analysed.

5.3.1 Sharing learning aims

Both teachers took students' age into consideration in designing learning aims. LT said "*senior pupils may find it [the aims] clearly, the younger ones may not be interested in these theoretical things*", because the younger students prefer physical faculties such as "*vision and hearing*". RT also agreed that "*when you bring up the concept of teaching aims, a child who is about 10 years old may not be able to understand it*". She also suggested a solution -

"*detailing*" the aims to make them understandable to young students. In RT's mind, the learning aims in a lesson were *"the aspects we can make better after we learned lesson*".

It was important to share learning aims with students. According to RT, the aims were like reminders for students that could give them directions and help them get prepared, like a "*lighthouse in the sea*". She believed that "*preparedness ensures success, unpreparedness spells failure*". The aims could quickly "*draw their attention consciously or unconsciously*" when there were 7-8 lessons every day. LT also mentioned that learning aims could serve as the criteria in self-assessment.

Learning aims were important to teachers too. Even though RT thought "*education was a traceless long process, cannot be quantified [by regulations like aims]*", she still believed that "*a lesson cannot be called 'a lesson' if the teacher did not have the aims*".

Lastly, both teachers agreed that SLA was helpful for students to develop "*learning habits*". RT emphasised "*confidence*" "*accumulation*" and "*lifelong*":

> *RT: Everyone needs confidence, but every piece of success or confidence was from the accumulation of little things… if a small aim was achieved, children will have a little of confidence. If we build up small aims, I believe the students will turn small aims into big aims, and little confidence into huge confidence… make sure that they keep the interests, step forward to the aims, I think this is the power, which can benefit [the students] for a lifelong time.*

That was why RT insisted on detailing the learning aims, not only because of the students' age, but also because of the aims' value in building up confidence and developing learning habits. Students proved that teachers' aims were clear enough to understand (see 5.2.1). It was only the level of difficulty in the aims that needed adjustment. Both the teachers and students acknowledged the importance of the sense of direction provided by SLA. Teachers also need aims to guide their own teaching.

5.3.2 Questioning

LT suggested that there were different purposes for each type of questioning. She thought that close-ended questioning focused on memorised knowledge, such as correcting "*grammar mistakes*", or "*capturing the details in the story*"; whereas open-ended questioning focused on cognitive-linguistic development, such as understanding a story or "*adding an ending for the story*". RT proposed that "*close-ended questioning was the foundation of open-ended questioning*". After confirming understanding of basic information with close-ended questioning, open-ended questioning could achieve further goals:

> *RT: open-ended questioning could broaden children's divergent thinking. It is helpful to the connection of children's knowledge, as well as re-construction. Actually, the clearest point of view from Constructivism is the re-construction of old and new knowledge, bring them into one.*

Both teachers agreed that they prefer open-ended questioning. As mentioned in the quotation above, RT believed that open-ended questioning could benefit "*students' thinking*

ability" by "*exploring in the practice*", and "*it was good for students to connect and reconstruct knowledge*". RT believed that teachers should ask more about 'what' or 'why', regardless of the problem with open-ended questioning that it can be time-consuming. LT found such questioning could "*light the fire of passion*", *because* "*more people can participate in the lesson*" and "*students can show their own opinions"*, which were similar to the reasons that students preferred it.

As Figure 4.15 showed, more open-ended questions and fewer close-ended questions were used in both classes, proving that teachers did put their words into practice.

5.3.3 Feedback

Feedback was the most common and most often used FA strategy. RT saw true feedback as "*expressing one's own opinion about a certain behaviour, on the foundation of love to the children*". LT did not give a definition, but she said that ideal feedback was "*varied and changing*". She admitted that there were some problems with her feedback skills, such as too much repetition and being too general:

> *LT: Sometimes, the positive feedback I gave them [the students] was repeated, and I use them automatically, in other words, it cannot show the characteristics of each student… It would not just be "good" or "very good", I, myself, found it boring.*

The teachers also listed the types of feedback they used in their classes. Similar to what the students have found (see 5.2.3), non-verbal feedback played a significant role in FAR. It was not only the "*body language*", but also "*eye contact*" and "*silence*" that conveyed messages for teachers. RT also believed that feedback should be multi-functional. She illustrated this with an excellent example:

> *RT: I remembered that a child wrote "how many lesson", everyone was laughing at him/her, I said "thank you for contributing a mistake". This sentence actually included a lot. First, tell him/her that this was wrong, but we did not refute him/her. [Secondly,] I think it may not be his/her own mistake, it could be a mistake that every student could make, just he/she let everyone know in his/her way. [Thirdly,] a feedback like this tells him/her the mistake, also protect her self-esteem.*

RT pointed out that the feedback should not refer to the child herself/himself, but instead only be aimed at the specific mistake or behaviour. This was what she insisted and emphasised:

> *RT: What I have been insisting is [feedback] should aim at the behaviour, rather than the individuals… You cannot say "you" are wrong, only "the word" was wrong, or "what you did" was wrong.*

Apart from the multi-functional feedback, RT also mentioned follow-up activities after feedback and a reward system which granted benefits to well-behaved students in FAR classes - for instance, immunity from homework for one occasion, or being a student teacher for a whole lesson. When engaged as teachers in this way, students could strengthen, expand and deepen their learning through teaching. According to students, there was also a reward system in FAL class but it involved credits. LT's credit system was more passive than RT's because students only received or lost stickers from teachers as

rewards or penalties, but the advantage was that everyone in the class could participate. Conversely, RT's system, in effect, only benefited the best students.

The teachers discussed the types of feedback which were not covered in the FA framework in this study, as well as the three kinds of verbal feedback with functions of confirmation, suggestion and correction.

LT said that she used confirmative feedback more than other forms, and this is evidenced in Figure 4.13. She believed that confirmative feedback was useful in motivating students and improving their achievement, because "*students could earn teachers' confirmation only by working hard*". Confirmative feedback also brought students "*the sense of achievement*". RT agreed that if *"[students] receive teacher's confirmation every day, then [the students] will become confident and successful*". In addition to simple confirmative utterances of one or two words (*e.g.* 'well done') used by both teachers, RT used more substantive, comment-like feedback, such as "*I like your pronunciation*" or "*louder please, what you said was right*".

> *RT: the content of assessment is really important, that is, the assessment itself, the way [of assessing], the behaviour, [are important], rather than a simple "well done". Assessment with content is, I tell the students that, today, I like your smile, like your pronunciation, like your well-behaved performance, like the way you look at me. All of these are assessments with content, they are not pallid or empty.*

Teachers used different strategies in situations where students did not offer correct answers. Both teachers would analyse the student's output first. If the student had already started the answer, LT would let the student finish. However, if the student was stalled or had nothing to say, LT would ask another student to answer, because she worried about the excessive time that might be spent by just one student on the question. However, she doubted this strategy herself, and it seemed that she was always in a hurry when waiting for students' answers.

LT chose different strategies depending on the time the student may take, whereas RT focused on the reasons why students could not answer. She identified three possibilities: the students did not know the answer, nervousness, and introverted personality. She would use suggestive feedback if the student could not answer the question, or when the student appeared "*confused and uncertain [about the answer]*". Her suggestive feedback took the form of rhetorical questions and analysis of the question for students. If the student was not confident, RT would comfort the student with encouragement first. Student interviewees from FAR, such as MGR and LBR, also mentioned these scenarios, and there is an illustrative quotation in 5.2.3.

Taking care with students' emotions was a dimension that LT did not seem to address. For instance, if a student could not answer, LT just turned to another student without giving any feedback to the first student. The reasons for the differences between the two teachers could be their work experience, their personalities, their competence of English language, and their underlying beliefs about education. RT may see more possibilities in students due to her rich teaching experience and demonstrable affection for students. Also, her corrective feedback was arguably more powerful in the context of her style of teaching. If a teacher is usually gentle and warm with the children, perhaps the children place importance

on that teacher's assessment and opinions of them, and might be more motivated to accept corrections of their mistakes.

What is more, as RT mentioned earlier, feedback should have substantive content. It is important to give the "explanation" immediately, or provide other acceptable behaviour, because it is only when children know "why this behaviour should not be done" and know "what is acceptable or correct", that they can improve their achievement. This is why RT emphasised that the corrective feedback should be "*acceptable and understandable*", and why student interviewees mentioned that suggestive feedback could help them "*remember how you learned it*" and achieve more.

5.3.4 Peer-assessment

Peer-assessment was not often used in FAL. LT admitted that she would "*use it once or twice in a lesson*". However, according to her description, she mixed up the concepts of peer-assessment and group discussion. She would ask students to discuss the "*controversial questions*" in groups, but she found it difficult to manage the class when students started to talk. It was "*time consuming*" and would "*interrupt the flow of teaching*". Also, she had a relatively aloof teaching style, which perhaps did not help her to work closely with students. These were the reasons why she did not use peer-assessment very often. Clearly, LT did not distinguish peer-assessment from group discussion, and she was not entirely ready and willing to relinquish any control to students during activities.

Both teachers realised that "*primary school students were good at pointing out others' mistakes*", because they were "*more innocent than grown-ups*". Due to this fact, LT found that "*positive peer-assessment was not as obvious as negative peer-assessment*".

On the other hand, RT had a deeper understanding of peer-assessment. This was conveyed firstly by her observation that students like to correct others because they were confident in their knowledge. They will use the 'right' knowledge to correct others or confirm the commonly held 'right' knowledge in peer-assessment. However, their so-called 'right' knowledge was not always correct. Because of China's One-child Policy, the attention from the whole family is on one person, and children nowadays are more individualised than before (Cai, Kwan, & Sedikides, 2012). Sometimes they might seem to be over-confident.

Secondly, even if there was an instance of negative peer-assessment, with help from the teacher, it could be turned into a good educational opportunity. She gave an example:

> *RT: I think that everyone, I believe that "Man's nature at birth is good" … But I need to tell you [the students who give negative peer-assessment], you should be tolerant of others' mistakes… Only with tolerance, children could reconcile themselves to life in the future, only by doing so, I think [they] could go further.*

Thus, RT's lessons were not only about English learning, but also about building character and understanding how best to live one's life.

Thirdly, peer-assessment would have positive impact only if students see themselves as equals. *"It should not be commanding, it should be equal [between students]"*. Attitudes of

superiority or condescension will make peer-assessment unacceptable. RT also listed three prerequisites of suggestive peer-assessment: a) the suggestions are possible and realistic to do; (b) the recipient must be willing to accept; (c) the recipient must be willing to put them into practice. A sense of equality makes the suggestive peer-assessment acceptable, and the knowledge from students and help from teachers make the peer-assessment do-able and practical. For the last prerequisite, RT also believed that:

> *RT: The children who are willing to give suggestions to others, are actually the students who are sensitive and willing to learn. They actually will correct their own mistakes. Also, if the students are willing to listen to others in the learning process, they would definitely work hard to present themselves, then, in the learning process, they also could adjust their attitude toward knowledge and the results of learning.*

Therefore, as a last observation, peer-assessment was closely related to self-assessment, which is a learning habit that can benefit students in the long term.

5.3.5 Self-assessment

Both teachers agreed that self-assessment was a learning habit. RT said that "*the development of the ability of self-assessment needs a very long process*". LT also saw self-assessment as important, yet she claimed that she did not use it. She mentioned the 'ticking time' activity instead.

> *LT: Ticking time just appeared in the textbook, so this module was new. There was no self-assessment before… [Students] with good learning habits already, they naturally have better motivation, some other children will only treat [ticking time] casually.*

However, to the researcher's mind, the scant use of self-assessment was not because of students' motivation to learn. As LT mentioned, self-assessment as a strategy was new to her as the teacher as well as to the students. Students may have felt unfamiliar with this form of assessment. They were still used to the traditional ways of learning, which centre on accepting only what teachers say. Therefore, it leaves the problem to the teacher: how to build the bridge for students between the traditional way of learning and the new way which emphasises learner autonomy. Students may need more scaffolding and help from teachers to learn how to use self-assessment well. This might especially be the case in schools in a second-tier city with a less advanced education system. The situation in big cities might be better, where there are more innovative ideas about evidence-based teaching and the latest learning methods are put into practice.

The help needed from teachers was not only about how to assess oneself, but also about how to do that assessment as critical as possible. LT mentioned the phenomenon that "*when students assess themselves, they are always inaccurate*". RT analysed the reasons.

> *RT: Because they are children after all, [they] could never see themselves in the round… Because they are not self-disciplined, in other words, their consciousness cannot reach the expectation of grown-ups.*

The consciousness that RT referred to, from the researcher's point of view, relates to the ability to be objective. Therefore, a lack of objectivity could lead students either to overrate or underrate themselves. RT took two examples: on one hand, the only child (*i.e.* one without siblings) may overrate themselves because "*they think they are the best*"; on the

other hand, the left-behind children[2] may underestimate themselves because "*they are not confident enough*". Therefore, "*teachers' guidance is the most important thing, only teachers will tell them objectively that what kind of assessment is suitable*". If the students underrate themselves, RT suggested two methods to address the problem: success experience, and attribution of success or failure to the behaviour of learning rather than the ability to learn. Young students cannot yet see the whole picture, so teachers can "*teach them how to assess themselves, then teach them how to assess themselves objectively*".

From RT's interview, it was obvious that she had already noticed the impact and problem of self-assessment, as did LT. Moreover, RT dug deeper into the issue, analysed the reasons, and found possible solutions. She believed that teachers also need self-assessment, to develop their practice.

That was self-assessment from teachers' perspectives. Taking their students' standpoint, teachers believed that self-assessment could improve pupils' achievement by helping them "*reflect upon one's own knowledge*" and working hard. It was easier said than done, but "*everything will be better only if we make great efforts*". Positive self-assessment would encourage students to improve themselves, but more importantly, being critical entailed "*the growth [of students] in boundless pain and boundless happiness*".

In summary, self-assessment, as a mode of formal assessment, was relatively new to schools in China. Its impact on motivation and achievement was positive. Some teachers, like RT, have already studied it and applied it in their teaching, while other teachers, like LT, were not accustomed to this new form of assessment. Where students used self-assessment, it was spontaneous, not dictated by the teachers. Teachers need to help students build a ladder from inaccurate judgements to more 'objective' self-assessment.

5.4 Summary of findings

In this chapter, I have discussed the findings from the qualitative data which emerged from the fieldwork interviews on FA with students and teachers. In the order of FA strategies, students' and teachers' interviews were quoted and analysed, to answer the research questions: what the students' perceptions of FA are and what the teachers' perceptions of FA are.

Students' perspectives on FA were complex. Different students had different attitudes. In general, students preferred positive or neutral assessment, such as sharing learning aims, open-ended questioning, confirmative and suggestive feedback, positive peer-assessment and self-assessment. However, there were some students who liked close-ended questioning because they found it easy, and some students liked negative peer-assessment because it could stimulate their drive to compete.

Both teachers could see the positive impact of FA strategies. However, there were gaps between theory and practice. Although the teachers had observed the phenomenon of FA, the rationale behind it and the improvements resulting from it were not studied in detail. LT

[2] The left-behind children refer to those who are left in rural regions while their parents work in urban areas.

seemed to have a general impression of FA, while RT understood the strategies better in terms of 'why' and 'how'. With more professional development and study of FA, teachers can narrow the gaps between theory and practice with their knowledge, experience and passion.

Sharing learning aims was a light 'starter' activity. Students liked simple and clear targets, which could be used as a guidance in self-assessment later. It brought students a taste of the upcoming lesson which could arouse their curiosity, as well as give them a sense of direction which could also reduce their uncertainty. To teachers, sharing learning aims was a "*lighthouse*" to show the direction of learning to students as well as teachers.

Questioning and feedback were the most often used strategies in the class. The strategies in questioning and feedback were like 'five main courses' cooked up by teachers. To teachers, questioning helped students to enhance existing knowledge and build new knowledge; feedback played the most important and complex role in FA, using different functions and forms, but with the shared requirement of substantive content.

Open-ended questioning, confirmative feedback and suggestive feedback were more popular than close-ended questioning and corrective feedback. Compared to the close-ended questioning, open-ended questioning could open students' minds because it encouraged them to express themselves freely and formulate answers in their own way. Open questions might be challenging, but HA students enjoyed them. However, for some other students, close-ended questioning was preferred due to its simplicity. Regardless of preference, most of the interviewees believed that open-ended questioning could improve their English because of the exposure to a variety of answers.

Personally, I believed that confirmative feedback was the sweetest, most appealing 'dish' in this FA 'fine dining' experience. Many students showed their strong preference for this type of feedback. Students can build confidence from confirmative feedback and feel encouraged by it. Nevertheless, its impact on achievement may be smaller than suggestive feedback. Suggestive feedback therefore might be regarded as the 'healthiest' main course. The main benefit brought to students by suggestive feedback was the improvement in their thinking capabilities, whereas corrective feedback may not necessarily stimulate independent thinking and might even discourage it. In other words, with corrective feedback, many students found themselves potentially becoming dependent on teachers' input. Even though they were only young children, the students' indicated a strong desire to be treated as independent persons and independent thinkers.

There was little peer-assessment and self-assessment used in traditional classes. Teachers believed that these were novel strategies which could lead students to become autonomous learners who could perform objective and critical reflection. These methods were like freshly produced 'desserts' which could be brought to the table whenever customers wanted them. Here, the students were the customers, as well as the pâtissiers, but they were only apprentices in that craft. They wished to take control of their own 'cooking' - their learning in this case. Therefore, to ensure the quality, students needed more clear, simple and practical instructions from the chef (teacher) or the recipe book (criterion-referenced regulations) to guide them to assess each other or themselves.

Most of interviewees stressed the benefits of negative peer-assessment in increasing motivation, even if it was not very welcome. Even though negative peer-assessment sometimes could be unpleasant, it would stimulate students' desire (especially LA students) to prove themselves to others. HA and MA students could see new aims in negative peer-assessment and could seek suggestions in it. With negative peer-assessment, students would have a deeper impression of the knowledge they were learning. Meanwhile, it was mainly HA and MA students who thought positive peer-assessment could benefit achievement. This might be because LA students did not experience much positive peer-assessment before. More attention needs to be paid to the LA group, as they were a vulnerable group in peer-assessment.

Self-assessment was clearly advocated by the educational authorities, as self-assessment forms were attached to each unit in the textbook. However, many interviewees claimed that there were no standards or criteria to use when they tried to assess themselves or their peers. Therefore boys, in particular, tended not to use self-assessment in reality. On the other hand, almost all the interviewees believed that self-assessment was helpful, because it could not only highlight their achievements and correct their mistakes, but it helped them know themselves better and reassure themselves. Hence, the gap between what was intended and what happened in reality begged the question - what was missing? To reiterate the findings above, the answer appears to be: clear, simple, practical instructions and criterion-referenced regulations.

Of course, many delicious menu items were not scrutinised in this *haute cuisine*, such as non-verbal feedback including gesture, eye contact and facial expression, and vital ingredients such as teachers' attitudes towards the students in class and students' personalities. However, these could be the subject of work in the future.

Levels of confidence in classroom practice caused by variations in teaching experience may be another reason that differs the two teachers. LT, as a new teacher, focused mainly on the forms of teaching delivery. For several reasons, for example the 'survival' dilemma she mentioned, she was not able to use FA well. FA was rigid and repetitive in LT's class, reflecting a one-size-fits-all approach to assessment and less enthusiasm on the part of the teacher and the students.

In contrast to LT's uniform use of FA, the successful use of FA must have a specific, tailored fit like a shoe. RT was not only able to use suitable strategies in different situations, she also wanted to pass on her philosophy and beliefs to students:

> *RT: I think whenever it is, what I want to teach the most is, wherever you go, wherever you are, only you would love yourself… Sometimes, I really want to say that it is impossible to tell children some big theories, but as a teacher, we take the responsibility of teaching and nurturing. Teaching is only about the delivery of knowledge, but nurturing tells people how to be a person, nothing is important in the world, it is important to let the students know that, in the world, the world could abandon you, only you cannot abandon yourself, this is the most correct self-assessment.*

Her ideal education involved a holistic approach to nurture a person by teaching the principles of independence and self-reliance. With such a clear goal, she insisted that FA strategies should help students build effective learning habits which could benefit them all their lives.

This chapter has analysed the research findings. The Discussion chapter (next) will compare the results of the current study with findings in other published research, discuss the evidence supporting the findings, and consider the new insights contributed by this study.

Chapter 6 Discussion

6.1 Introduction

After analysing both the quantitative and qualitative data, several findings have been identified with solid evidence. In this section, on the basis of the findings in Chapters 4 and 5, I will answer the five research questions posed in the Chapters 1 and 3, explain how the results support the answers, and how the answers fit in with what was already known about the research problem being investigated. I will then discuss how the findings add new understanding or insights to the field.

In social sciences, one of the purposes of research is to discover new information, not to prove a hypothesis as in pure sciences (Wellington & Szczerbinski, 2007). Therefore, all the possible explanations for the study's results are be carefully considered, regardless of whether they fit the researcher's prior assumptions and biases or not. In particular, when discussing the actions in the classroom, a very complex environment, and pupils' psychological development or changes, an even more complex and not yet fully discovered area in science, the connections between the results and explanations will be discussed here no matter how strong or how weak those connections appear to be (Cohen, Manion, & Morrison, 2011).

Research questions were formulated in 3.2.1 on the basis of Literature review. Three of the research questions were about the relationship between FA, motivation and achievement, and two questions were about the impact of FA on motivation and achievement from students' and teachers' perspectives. As mentioned in Methodology and Methods, the quantitative and qualitative data are complementary. Therefore, in the following sections, I will use the interrelationship between FA, motivation and achievement as scaffolding, to discuss not only the facts of 'what' from quantitative findings, but also the reasons 'why' from qualitative findings and previous literature. This approach also reflects the Positivist and Interpretivist epistemology explained in 3.3 Philosophical foundations.

6.2 FA strategies

Formative assessment is one of the major influences on student motivation (Crooks, 2001). The quantitative and qualitative findings summarised in 4.9 and 5.4 confirmed the statistically strong relationship between FA and motivation, FA and achievement, which suggested that FA has a positive impact on students' motivation and achievement in general. This finding accords with the work of Afitska's (2014b) where both teachers and students showed a positive attitude to FA strategies. Therefore, even though the use of FA may not always lead to improvements in achievement, there may be a positive impact on students' attitude to FA (Lam, 2016). Afitska (2014b) also suggested that FA strategies, namely, feedback, questioning, peer-assessment and self-assessment, have a positive effect on linguistic development. Some empirical projects have also proved that FA can improve students' achievement (Andersson & Palm, 2017; Carrillo-de-la-Peña, et al., 2009; Wiliam, Lee, Harrison, & Black, 2004).

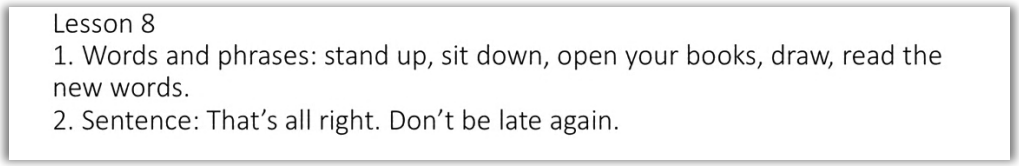
The quantitative findings in the comparison across groups in this study revealed that there were differences between groups regarding the impact of each FA strategy on students' motivation and achievement. Although students firmly believed that FA was very helpful to their achievement, the impact size was not large in the quantitative results, except for one sub-group: LA girls.

In the following sections, the impact of FA on motivation and achievement will be discussed in respect of each FA strategy with students' and teachers' comments. The similarities, contradictions, differences were highlighted in combination of the findings.

6.2.1 Sharing learning aims

The teachers were sharing the learning aims (SLA), but the form of it was simple and fixed. All the student interviewees were very unfamiliar with the concept of sharing learning aims. MBL did not even understand it after my explanation. It is understandable that SLA as a learning concept is not well-known among primary school students as they are not education scholars. Therefore, they provided some examples that they are familiar with. To the interviewees, the forms of the aims were the target words/phrases or which part in the textbook to learn, even what to do in the lesson. When LT herself was talking about SLA, she kept referencing it to self-assessment 'ticking time' in the textbook. It seems that LT sees SLA as the same as self-assessment criteria. With two years of teaching experience, LT linked SLA closely to textbooks, but had started to relate SLA to learners' needs, an encouraging development for a novice teacher.

During the fieldwork, a common occurrence was noticed in both school campuses: before each lesson, there would always be a paper memorandum about the target words and speech patterns handed out to each student (see an example in the picture below). Because the two teachers used the same preparation material, all the students had the same memorandum. It was possible that students thought these were the learning aims.


Lesson 8
1. Words and phrases: stand up, sit down, open your books, draw, read the new words.
2. Sentence: That's all right. Don't be late again.

The memorandum merely listed the words and sentence, which was useful for students to recite, as HBL and MGL mentioned. However, there were no clear criteria about the words, phrases and sentences. For example, students did not know whether they needed to recite the all the phrases or just as much as they can; whether they need to know how to read the phrases or how to use the phrases; whether they need to know how to answer questions with the sentences or apply them into daily life. From the interviews, most of the students believed that the aims were used to recite the words, phrases and sentences. However, it was not only the language knowledge students needed to learn in the class, but also other aspects of language learning, such as the intercultural knowledge. Harris (2007) also suggested that aims emphasising activities are less helpful. Unfortunately, these other aspects were not shown in any of the memoranda.

The difficulty of the aims and the frequency of SLA use can affect students' motivation. FAL students were showing stronger correlation between SLA and motivation than FAR students (Appendix, Table 5). However, after 10 weeks, the correlation in FAL was even weaker than FAR's. It may be because LT shared aims that were beyond the students' ability, compared to students from FAR, given that more students from FAL found the aims harder. FAR students commented that RT can, as HGR said, "*connect the aims to lessons*". RT had more teaching experience and life experience, which may have helped her to apply the textbook more flexibly and see learning aims from a relatively macro perspective. Accordingly, it may be easier for her to set learning aims beyond the textbook. The examples RT gave focused on not only the learning activities, such as finishing a section in the textbook, but also mastery of knowledge, like retelling the story. In terms of SLA utility, one of the differences between RT and LT was the frequency of sharing the aims (Figure 4.13). All students showed similar motivation in response to SLA (Figure 4.18), but the correlation was weaker in FAL. Hence, the frequency of SLA use may lead to the changes in the SLA-motivation relationship.

The level of difficulty of the learning aims shared by teachers during the fieldwork was not suitable for everyone. HA student interviewees all found the aims easy, but MA and LA found them sometimes difficult. To students, achievable aims can help students to build confidence and subsequently achieve even greater aims. HA students believed that aims can provide them with "*a sense of direction*", making it "*easier to achieve knowledge*". RT argued that aims should be achievable so that students can grow more confident. She also believed that due to the intensive nature of the school day in China, SLA can signpost the lesson like a "*lighthouse*". Meanwhile, LT emphasised the aims should suit students' age, that the cartoon format is better for younger learners, and theoretical aims are better for students in senior grades.

Students liked to work harder if the teacher gave them more detailed aims. The aims HA mentioned were similar to what RT suggested as "*small aims*". As quoted in 5.3.1, RT was committed to the idea that "*quantitative accumulation leads to qualitative transformation*". By small aims, I believe RT referred to aims that are suitable to students with different achievement levels. This may explain why SLA was not related to HA's motivation in the post-test, but was strongly related to LA's motivation at both times (Appendix, Table 4). All the students were looking forward to new lessons with new aims, but aims without a hierarchy of difficulty may cause HA students to lose interest, but may spur LA students to think harder. Furthermore, the teachers focused on the "*sense of direction*" in SLA, but did not set landmarks in different stages. Students in HA may still find their way through the learning quite easily, but those in LA may actually need more help.

Most students saw learning aims as the teachers' guide to their study, whereas highly self-motivated students had their own aims. Since the aims in the class were clear, most students were happy about the use of this strategy. Some students did not need aims shared by teachers in their study, because they had their own aims all the time. As HBL said, English is "*seen as a part of his life, you are willing to learn English, rather than forced by the teacher*". He had an extremely positive attitude towards language and language learning. Therefore, in the goal setting phase in Dörnyei and Ottó's (1998) framework (Table 1 in Figure 2.8), this student had demonstrated autonomy, which enabled him to monitor his own learning progress and created more accountability via SLA (Tell, Bodone, & Addie,

2000). Similar to what RT suggested that the aims she wanted to share were not for just one lesson, they applied continuously throughout English study, even addressing students' character development. However, students like HBL are in the minority. Other students still need the teacher to share the learning aims with them. RT set aims for the purpose of giving direction, and LT focused on the form of the aims. RT spoke of why we set aims, how to set aims, and what is in the aims. Although RT and LT had different understanding of SLA, they showed that they were reflecting on their use of the strategy, and critical reflection is an important aspect of professional development and learning. More importantly, as setting learning aims is no longer the sole prerogative of the teacher, involving students in the goal-setting process might maximise the effect of SLA (Curry, Mwavita, Holter, & Harris, 2016).

Lessons being interesting, and this being conveyed in the aims, is one of the key elements in motivating students and improving their achievement. Girls' SLA was moderately related to their motivation during the fieldwork, but for boys this correlation declined sharply from strong to weak (Appendix, Table 6). According to the interviews, girls were curious about what to learn in the lesson via SLA, whether the lesson is interesting or not, and they found aims easier compared to boys who kept an uncertain or negative attitude towards SLA. This explains why SLA was closely related to motivation factor *Language* among girls. On the other hand, boys were less able to adapt their motivation level compared to girls (Oga-Baldwin, Nakata, Parker, & Ryan, 2017), so it may be relatively difficult for them to adjust their attitude if they hear 'not so interesting' learning aims.

According to the quantitative results, only the LA group had a moderate correlation between their achievement and SLA (Appendix, Table 8). The impact of SLA was not statistically significant in HA and MA groups. LA interviewees all believed that SLA can improve their achievement, but only one explained the reason. It was because "*what teacher said can let us know whether this lesson is interesting or not*". The answer was related to her motive but not how or why SLA was helpful. From her response, it was apparent that motive foreshadows the efforts that students may put into their study. This was the reason that the impact of FA on students' motivation was analysed and discussed first. In other words, motivation was a prerequisite to making efforts, and making efforts was a prerequisite to making progress in academic achievement.

6.2.2 Questioning

Most of the student interviewees preferred open-ended questioning, so did the teachers. Although LT used more close-ended questions and RT used more open-ended questions, according to the data gathered from students (Figure 4.13), both teachers claimed in their interviews that they preferred open-ended questioning. Flexibility in the open-ended questioning may motivate students and improve their achievement. Open-ended questioning enabled more opportunities for different answers by students, making this a dynamic strategy which brought new information into the class.

All FAR students but one preferred open-ended questioning, and half of the FAL students preferred close-ended questioning. This matched the outcome shown in in Figure 4.13 where FAR used more open-ended questioning and FAL used more close-ended ones. Firstly, it is possible that these phenomena were caused by the use of different questioning.

Students were exposed to different amounts of the two kinds of questions, and as a result they may like the type that they are familiar with. Therefore, FAL students preferred close-ended questioning, while FAR students preferred open-ended questioning. Secondly, FAR's questioning-motivation relationships were increasing while FAL's were declining during the fieldwork (Appendix, Table 5). Many students emphasised "*express freely*" and "*varied answers*" in open-ended questioning as the reasons for preferring it. The flexibility in open-ended questioning suited the FA environment in FAR which encouraged students to explore beyond the textbook. It enhanced the students' motivation more and more during the fieldwork.

Self-efficacy was one of the reasons that students adopted approach or avoidance tactics towards questioning. Most of the interviewees, HA students in particular, preferred open-ended questioning because it was more interesting than close-ended questioning due to its varied answers. Motivationally, students felt proud if they could come up with different opinions. They were aware of their strengths when their performance was rewarded with a sense of achievement (Table 4 in Figure 2.8 and Self-concept in Table 2.10). In terms of achievement, the open-end questions could help students to improve their speaking skills. They have more opportunities to practise the language knowledge they have learned, and express themselves freely, rather than replying with dichotomous answers. In their minds, the more you speak, the higher the score you might get, and the better your communication skills will become.

However, there were some interviewees who preferred close-ended questioning, because they believed such questions were easy to answer, or even easy to guess the answers. HBL, for example, said that "*the answer in open-ended questions is too long, I cannot organise my language well*". In fact, he was one of the top students in his class. However, perhaps due to his fear of failure (Table 2 in Figure 2.8 and Achievement Theory), he was more inclined to answer close-ended questions to avoid being wrong (Avoidance State of Mastery goal in Table 2.7).

Although teachers and students had preference between open-ended and close-ended questioning, teachers were acknowledged that both types of questioning were equally important. They agreed that different question types need to be used in different situations: for example, close-ended questions to check the linguistic facts, open-ended questions to provoke thought or opinions. Although Harris (2007) claimed that questioning is effective in creating learning not just testing it, the importance of basic memorial knowledge in class cannot be ignored.

In the real classes, students did not have the same chances to answer questions. LA students were influenced by questioning but did not receive enough attention from the teachers; in contrast, HA may not be as motivated as they thought by questioning. LA students' motivation was strongly influenced by both types of questioning (Appendix, Table 4). Both kinds seemed equally important to LA students because they did not indicate a preference (Figure 4.17). As RT and LT commented, close-ended questioning was used to 'check' the basic knowledge that has been learned already, and open-ended questioning was used to 'explore' the new knowledge. LA students needed the external confirmation, as they may be unsure of themselves. Meanwhile, they also have the willingness to explore

and reconstruct the knowledge, like MA and HA. Unfortunately, they received the lowest amount of questioning in the class (Figure 4.15). On the contrary, although HA students claimed that they preferred and enjoyed the challenges in open-ended questioning, their motivation had no relationship with open-ended questioning, and had a merely moderate correlation with close-ended questioning in the post-test. It is natural that most people prefer to engage in the topics within their abilities and stay in their comfort zone, let alone a special group such as children. The LA group was like a drier sponge compare to HA - they would like to absorb as much as they can. HA students may find that the strategies used to motivate and help them to achieve more were no longer interesting. Hence, open-ended questioning had little impact on HA's motivation.

The amount of questioning did not necessarily have impact on students' achievement, but how the students see questioning could. Interestingly, close-ended questioning was used less in FAR, but had greater (but still weak) impact in this environment than in FAL; while open-ended questioning, which was used less in FAL, had a greater (but moderate) impact in FAL than in FAR (Figure 4.13 and Appendix, Table 9). It was true that RT created an FA-rich environment, where more open-ended questioning was used. What RT wanted was to let students "*explore new ideas*". Open-ended questioning did not seek quick answers such as memorised facts, but instead sought something beyond the superficial level (Black, 2003; Liu & Xu, 2017). However, when it comes to the students in this study, the function of open-ended questioning was, in effect, "*maybe one answer will not be on the exam paper, but other answers could be*". The exam-oriented learning style was still deeply rooted in students' minds. Therefore, it was not only teachers' beliefs about teaching and learning that needed to be changed, but those of students too.

The reasons for students' preference were different in boys and girls. Girls preferred varied answers, and boys preferred simple and direct answers. All girls and boys from LA preferred open-ended questioning, and they commented that it provided more possibilities and students can learn from each other. This was echoed in the discussion of FA and achievement that most of the girls and half of the boys agreed that open-end questions were more helpful in language achievement. This is reflected in the relationship between questioning and achievement too (Appendix, Table 10): close-ended questioning had a greater impact on boys' achievement; open-ended questioning had a greater impact on girls' achievement.

For boys, close-ended questioning was easier to answer because they "*cannot organise language very well for long answers*". The strength of the correlation between boys' motivation and close-ended questioning was stronger than between their motivation and open-ended questioning. It was the opposite trend for girls (Appendix, Table 6). Some girls called close-ended questioning "*dead*" and "*boring*", other interviewees believed that open-end questioning had varied answers, and students had more opportunities to express themselves. The gender difference in the questioning-motivation relationship was probably related to the nature of the questioning and the gender characteristics that students showed in the fieldwork. According to the interviews, girls showed more interactive tendencies, while boys showed more analytical tendencies. This fits with the body of the research literature (*e.g.* Baron-Cohen, 2004) which argues that girls are better at empathy,

while boys are better at analysis. Although gender characteristics can be affected by societal factors, the difference between them is stable (Ellis, 2008; Halpern, et al., 2007).

Teachers' roles involve not only teaching practice. They must engage in continuous professional development and study of the latest innovations and theories and, importantly, have the ability to put the theories into practice. Most of RT's perceptions were based on her strong theoretical background. RT cited education theories such as Constructivism and Discovery Learning and explained how open-ended questioning is useful. This stance may be because educational reform in China was underpinned by constructivist theories (Tan, 2017). On the other hand, LT relied on her intuition and observation. Although she had recently qualified and theoretical knowledge is very important in initial teacher education, LT did not apply her education theories in her practice. In contrast, RT had the advantage of rich experience and skills in both theory and practice.

Time management is vital to teachers, as they need to engage more students and avoid spending too much time on a few (Harris, 2007). In this study, waiting time concerned teachers, during questioning as well as feedback. Teachers normally allow less than one second to pass after asking a question before posing another one if nobody is going to answer (Rowe, 1974). LT was also troubled by waiting time: "*I was very anxious if the student cannot answer a single work in a few seconds*". Depending on the time taken by students, she saw two possible scenarios: a) the student did not know how to answer; b) the student was not listening. If the latter, she would use her classroom management skills; if the former, she would give suggestive feedback or let the student wait or sit down. It was clear that a well-organised class was needed to implement FA properly. Therefore, LT's concerns about classroom management were understandable for a novice teacher. In contrast, RT's extensive experience managing classes made it easier for her to use questioning with skill and ease.

6.2.3 Feedback

There was no doubt about students' preference for confirmative feedback. All the students liked it because they may feel happy and encouraged with more confidence after their efforts have been recognised by teachers. Young learners needed positive 'energy' in feedback. Interviewees all wanted to be rewarded when the teacher gave feedback, whether it was confirmation or advice on constructing the way forward. This reflects the literature, where teachers' support has been related to higher student motivation (Carreira, Ozaki, & Maeda, 2013). Students expressed their negative feelings about discouraging conversations with teachers, and positive feelings about being supported or encouraged by teachers and peers (Xiao & Carless, 2013). If the teacher behaved in negative ways with students, as MGL said, "*it feels that the teacher gave up on you*". However, perhaps because it was the most often used FA strategy (Figure 4.11), students may take it for granted and had less reflection on it.

Although it has been claimed in other studies that students hardly do any independent thinking and become 'fact memorising machines' (Chen, 2015), students in this study showed that they already realised the importance of independent thinking in learning. All the interviewees preferred either suggestive feedback or confirmative feedback. In the LA

group, all the interviewees preferred suggestive feedback. LA students were eager to receive suggestive feedback, albeit they received much less suggestive feedback than the other two groups (Figure 4.12). The reason students liked suggestive feedback was that it enabled students to "*think on their own*". The confidence that inspired by confirmative and suggestive feedback can push student towards independence (Jones, 2010). This phrase appeared repeatedly during the interviews. Even though all the interviewees were at young age, the awareness of the importance of independent thinking was rooted in their thoughts.

Confirmative feedback can improve students' achievement in the short term, while suggestive feedback may have a positive impact in the longer term. Only confirmative feedback was moderately related to FAR's (Appendix, Table 9) and boys' (Appendix, Table 10) achievement. Students liked the independent thinking required in suggestive feedback, yet their achievement was actually improved by confirmative feedback. The recognition from the teacher may be the reason why FAR students and boys can improve their achievement. It was also possible that the short period of the study, just 10 weeks, was not long enough for the impact of suggestive feedback to emerge. Because confirmative feedback enhances the knowledge that students have mastered, and suggestive feedback helps them to construct new knowledge alongside previous learning, it will take time for students to digest and master the new information.

All students showed a weaker relationship between corrective feedback and their motivation in the post-test (Appendix, Table 4). The decrease in the use of corrective feedback and students' attitude towards it was considerable (Figure 4.15, Figure 4.20). Because the teacher would give out answers directly in corrective feedback, students are afraid that they will develop a pattern of waiting for teachers' answers if they are used to the routine of 'do not know how to answer' then 'teacher provides answers', and accordingly would be less likely to consider the questions independently themselves. Although all LA interviewees believed suggestive feedback was the most helpful feedback strategy, it was actually corrective feedback that had a moderate relationship with their achievement (Appendix, Table 8). No one in the interviews suggested that they were motivated by corrective feedback, or their achievement was improved by corrective feedback, probably because they may be misled and think they lack ability to study if they receive too much negative feedback (Black & Wiliam, 2004). However, corrective feedback was the most useful feedback strategy to improve LA students' achievement. Corrective feedback reveals teachers' emphasis on linguistic accuracy (Xiao & Carless, 2013). This may relate to the fact that LA students had relatively weaker basics in their knowledge. Correcting their mistakes could be the most direct and the most effective method to help these students to improve. The negative misconceptions about corrective feedback may lead to lower motivation derived from this strategy (Figure 4.17). However, corrective feedback itself was not enough, as time went by: the positive impact of corrective feedback on achievement may disappear if there are no follow-up activities (Afitska, 2015).

The rich FA environment would enhance the impact of suggestive feedback on students' motivation, but it is hard for new teachers to achieve this. Students had similar attitudes towards confirmative and suggestive feedback (Figure 4.18), but FAR students built a stronger correlation between suggestive feedback and motivation than FAL, as well as when compared to the pre-test (Appendix, Table 5). This may due to how much the strategy was

used in different FA environments. There was considerably more suggestive feedback in FAR than in FAL (Figure 4.13). A climate that encourages exploration and negotiation rather than simply forming judgements is more likely to promote learning (Harris, 2007). In interviews, LT humbly admitted that as a new teacher, her feedback was not good enough due to a lack of variety. She mainly gave short confirmative feedback. For novice teachers, the creation of a rich FA environment may be beyond the scope of their attention when they still struggle with classroom management.

Girls' empathy and boys' preference for straightforward input may be part of the reasons for their different preferences in feedback. As can be seen from the correlations between feedback and motivation (Appendix, Table 6), boys' interests initially lay in suggestive feedback, then changed to confirmative feedback in the post-test; meanwhile, girls' interests changed from corrective feedback to suggestive feedback. As girls mentioned in the interviews, they do not like corrective feedback, because they are concerned they would not think on their own once they pick up the habit of waiting for the answer from the teacher. What is more, they may feel that the teacher had given up on them if corrective feedback was given too easily. Figure 4.14 shows that girls received more confirmative and suggestive feedback than boys. It is possible that girls' empathy makes them particularly sensitive to these two types of feedback (Baron-Cohen, 2004), and they are more likely to be motivated by them. In the fieldwork interviews, girls showed that they were more cautious than boys. This characteristic may draw more teachers' attention and patience to them in class. Meanwhile, boys attempted to save their time and energy by avoiding complexity. Confirmative feedback can give them direct recognition for 'making progress in English'. It was not only the straightforward feedback that boys liked; they also valued shorter and simpler learning aims (MBL), questioning (HBR), peer-assessment (HBR / MBR).

Young learners may ignore the positive impact of undesirable feedback if they do not treat it critically. In the discussion above, it seems that the feedback students liked was different from the feedback that improved their achievement. It was true that students had negative feelings towards some feedback, such as corrective feedback, because it was hard for young learners to take correction as help instead of criticism. Feedback may be unappealing if it reminds them of their deficits. However, students, at least in this study, still held prejudice towards corrective feedback. If the teacher and students themselves let their prejudice deepen, it may demotivate them and subsequently lower students' achievement.

In the study, corrective feedback was used wisely. RT used mistakes as an opportunity to improve students' understanding. In the example illustrated in 5.3.3, RT skilfully turned a mistake into a good lesson for everyone in the class. This indicates that RT's feedback focused on the behaviour itself rather than the individuals. At the same time, she saved the student's *mianzi* which protected his/her self-esteem. These observations chime with Harris's (2007) work showing that wrong answers can bring new experience to students and promote learning.

Being critical was not only important to students when they received corrective feedback, but also important to teachers when they use confirmative feedback. In the interview, MGR raised another new perspective on confirmative feedback – that it should be used critically. She mentioned conceitedness of students and misleading self-assessment due to excessive

confirmative feedback (quoted in 5.2.3). It is hard to define 'excessive' here. When I started my PhD project, as mentioned in Chapter 1, there was always a question in my mind: can confirmative feedback be beneficial all the time or how much confirmative feedback will be enough? I had an idea about calculating the 'perfect' amount of each FA strategy for different students. I realised this goal was unrealistic after I stepped into the classroom. There were so many factors that needed to be considered, not to mention the unexpected matters that emerged. Hence, whether feedback is 'excessive' or not, needs teachers' judgement. There were some signs showing a tendency for excessive feedback, for example, as MGR mentioned, conceitedness in students. Teachers cannot give empty positive feedback, but substantive feedback that is suitable in the circumstances.

Teachers' feedback styles were different because their approach to teaching and learning was different. LT's priority was the content in the textbook, but RT also embedded life lessons in her teaching. In feedback, RT identified areas of achievement and gave suggestions on ways to improve, especially in terms of being a person in society. LT focused on linguistic tasks, on faults in interpretation of tasks, or on outcomes. Studies have shown that feedback will be invaluable if it is thoughtful, reflective, evokes thinking, explores understanding, and provides clear information on how to improve or reduce learning gaps (Black & Wiliam, 1998b; Nyquist, 2003). RT's suggestive feedback might be useful for capable learners who can reflect via teachers' feedback, whereas for students who need more detailed help with linguistic content, LT's feedback would be more suitable.

However, oral feedback was not enough for students anymore. Students from FAL complained that LT gave "*nothing else*" except simple confirmative feedback. It seemed that they were not satisfied with their teacher's feedback. HBL mentioned especially the comparison between LT and his English teacher from last year. The previous teacher always combined different forms of feedback together, which he liked very much. LT used fewer forms of feedback than RT. According to the classroom observations, RT used plenty of body language, especially hugs. HBR also mentioned that body language was often used in FAR, for reward and punishment. RT herself mentioned that she uses eye contact often too. As mentioned in Chapter 1, being in the primary school classroom in China is different from being in the British classroom. Students need stimulation during such a long day. Physical activity in the primary school classroom has been proven to positively impact students' academic achievement (Vazou, Gavrilou, Mamalaki, Papanastasiou, & Sioumala, 2012). Therefore, physical actions or exaggerated tones are effective ways to engage and stimulate students.

Teachers' follow-up activities were also varied. RT used more flexible activities, such as the 'little teacher' activity, which gave students more ownership and control of learning. In contrast, LT organised a credit-based system as a supplementary reward or punishment for pupils, but activities stayed under the control of the teacher. It has been argued that feedback is more effective without marks or grades, because marks often dominate students' thinking. Students will stop at the numbers, such as their scores in LT's credit system, not move forward to the written or verbal qualitative feedback (Crooks, 2001). However, LT's follow-up activities impacted on every student in the class, unlike RT's activities which benefited only a few students.

6.2.4 Peer-assessment

Being recognised by others was one of the biggest motives that LA students had. Hence the positive peer-assessment was very influential to them because it was encouraging after being recognised. Due to Confucianism's tenets, Chinese students tend to give positive peer-assessment to avoid tension and disagreement, meanwhile to save *mianzi* for others (Carson & Nelson, 1996; Cheng & Warren, 2005; Carless & Lam, 2014; Poole, 2016). They do not like negative peer-assessment because they believe sometimes it may hurt people's feelings, because it is too frank or candid. As mentioned in 3.7.1, children are special because there are many conflicts within them. They are frank, but at the same time they are sensitive. Especially for girls, they did not like negative peer-assessment and boys had no preference between positive and negative peer-assessment (Appendix, Table 6). They may want to protect their self-esteem, so they preferred positive peer-assessment. Girls worried about negative peer-assessment being "*too candid*" and "*awkward*", which may hurt their feelings, causing frustration or putting them in a bad mood. More girls showed fearfulness than boys. Young children are more sensitive compared to other language learners such as college students. Hence, seemingly minor influences may have a great impact on children. *Mianzi* then becomes a serious barrier to implementing FA in the Chinese school context, especially in classes with large numbers of pupils (Gu, 2014; Chen, May, Klenowski, & Kettle, 2014). When using peer-assessment, negative peer-assessment in particular, it is worth paying extra attention to children's emotions, not just language achievement.

There were some students who liked negative peer-assessment. Their attitudes were of two kinds: a) reflection: the negative peer assessment can point out their weaknesses, so they can work hard on these areas, and improve themselves; b) motivation: the more negative peer assessment they received, the stronger their willingness to prove their abilities or stop others from looking down on them. The first perspective can relate to the Attribution Theory of motivation. Students put information together until they find a reasonable explanation for the negative peer-assessment, then act appropriately to avoid the same failure (Heider, 1958; Weiner, 1972). The second perspective, regarding motivation, was mainly reported from lower achievers. It can relate to the self-worth element in Self-efficacy Theory and the Performance goal in Goal Orientation Theory. Students' judgements about their own self-efficacy were challenged by peers (Schunk & Pajares, 2009). They wanted to justify themselves by focusing on avoiding inferiority, and not looking stupid or dumb in comparison to others (Pintrich, 2000). Moreover, it was found that Chinese students were more likely to experience negative feelings, such as anger or shame, in learning because of the testing system, but the experience can activate greater achievement due to students' self-regulatory efforts (Chen & Brown, 2018).

LA students were more eager to prove themselves to others that they have the ability to learn than other two groups, therefore, negative peer-assessment was useful to them, but it may "*hurt one's pride*" as MA and HA explained. Negative peer-assessment was negatively related to LA's motivation in the beginning, but this turned into a positive relationship after 10 weeks (Appendix, Table 4). In the interviews, all the LA students believed that negative peer-assessment was helpful because they were willing to work harder to prove themselves to others. Even though boys feel embarrassment in negative peer-assessment, they believed that it can remind them of where they can improve, especially the simple forms of

assessment. But there were also some boys who did not like peer-assessment at all because they claimed to be shy, or believed that the standards differed from person to person. Boys showed less helpfulness than girls, which may relate to their characteristics of inhibition (Côté, Tremblay, Nagin, Zoccolillo, & Vitaro, 2002).

However, LA students did not receive enough peer-assessment in the classroom. The students in HA received and gave positive peer-assessment to each other. When they talked about examples of peer-assessment, it seemed all the examples were positive. This is perhaps due to their worries about the "*unpleasant effect"* that negative peer-assessment may bring to others. Both MA and HA preferred positive peer-assessment, but LA students did not show a preference for either kind of peer-assessment (Figure 4.12 and Figure 4.17). Meanwhile, LA students may not receive much peer-assessment at all, not to mention positive peer-assessment. When I asked LBL about the reactions to positive peer-assessment, he said "*I do not know the feelings*" because he had seldom received it before. The balance of peer-assessment between each group needs to be guided by the teachers.

Students had their own learning habits, which include assessing each other and themselves. It is in people's natures to form their own judgements about others. However, if there is no proper training, it is hard to turn this into effective peer-assessment or self-assessment in language learning (Wang, 2014; Zhao, 2014; Chen, May, Klenowski, & Kettle, 2014). The correlations between peer-assessment, self-assessment and students' motivation were very similar in FAR and FAL (Appendix, Table 5). The use of peer-assessment in FAR was slightly higher than it was in FAL, but self-assessment did not show any difference (Figure 4.13). According to the interview, three boys said that there was no peer-assessment, such as LBL, and the other two from FAR mentioned that if RT asked them to applaud, they would do it mechanically, but otherwise they did not assess each other. Their responses were backed up by the quantitative data in Figure 4.11 showing that peer-assessments, negative peer-assessment in particular, were the least used FA strategies in class. There was no classroom climate for formal peer-assessment in this study. The formal peer-assessment that did happen was mostly led by the teacher, and the students saw it as tokenism. MBL described himself as feeling lost in the peer-assessment when he was asked to do it. The students who are giving the peer-assessment may see it as tokenism, but those who are receiving it may still be positively affected by it.

The students' diffidence and bias may cause avoidance of peer-assessment. Many male interviewees did not like peer-assessment due to various reasons, such as their personalities and perceived unfairness in assessment. Of course, there were some students who insisted that they do not need the "*judgements from others*", such as two interviewees in MA group. This may be due to personality factors, such as embarrassment felt by students who were shy or had low self-esteem. MGR suggested that she can improve herself via peer-assessment, as well as help others. This may due to her supportive personality too. As MGR explained, students may feel awkward or even sick of peer-assessment, especially negative peer-assessment, because they cannot bear the embarrassment.

Secondly, it may be due to students' lack of confidence in their assessing ability. HBR did not like peer-assessment because he thought "*students' standards are varied, everyone should reach teacher's standard*". Compared to student assessment, they preferred the teacher as

the reliable source of feedback. Students do not see student assessment as 'real' assessment (Harris & Brown, 2013). Regardless of students' ability to assess each other, their belief had partially determined that they did not like peer-assessment. However, Harris (2007) and Black (2003) argued that students are more amenable to accepting criticism from peers than teachers who have power over them. It may be dialectical because it depends on how students see peer-assessment, as help or merely as criticism. In this case, the attitude of the assessment giver was vital to influence the receivers' perceptions. A peer-assessment with a friendly attitude may be more effective than with an indifferent attitude.

Lastly, personal relations can affect the peer-assessment. LGL complained that she did not trust peer-assessment to improve achievement because of "*the bias among the students*". For example, the relationship between students may affect the peer-assessment. HGR implied that she always received "*just so-so*" comments because of the uneasy relationship with her desk-mate. This phenomenon was also found in Chinese universities in other research studies (Xiao & Carless, 2013). Personal relations and the ability to peer-assess may affect the use of peer-assessment (Afitska, 2015).

According to the perspective of interviewees, positive peer-assessment can fill the gap between being motivated and making efforts. As HBL said, it "*pushes you to climb a higher mountain*". It seemed that the students cared more about their peers' attitude, than the contents in the assessment. For example, MGL mentioned that she can achieve more by correcting the mistakes pointed out by her classmates, which she referred to as positive peer-assessment. RT strongly argued that "*the domineering assessment is inappropriate*". She believed that peer-assessment should be for helping purposes, rather than showing one's superiority. In China, most of the children are the only child in their families. They become self-centred more easily and can stubbornly insist on their opinions, even if they are wrong. However, peer-assessment can be effective when it is appropriate and friendly, making the receivers willing to listen, and willing to take action in response.

The positive impact of peer-assessment existed, but it needed more time to affect students' achievement. Positive peer-assessment had significant but weak correlation with students' achievement (Appendix, Table 8, Table 9, Table 10). In other studies, peer-assessment was found to lead to greater self-regulation and achievement in students (Harris, Brown, & Harnett, 2015). Higher-achieving students tended to give more feedback in peer-assessment than in self-assessment. HGR gave an example of peer-assessment. Her classmates helped her with using 's' after the third person verb form. She was able to remember it, but in the exam, she forgot to add an 's'. After that, the rule was more deeply rooted in her memory than before. Although she failed in the exam question, she had already realised that peer-assessment was useful in uptake. Most students believed that it was negative peer-assessment that could help them more in their achievement. Although negative peer-assessment motivated students (mainly boys and LA students) to some extent, its impact on students' achievement in this study was small, and this may be due to the short timeframe.

The findings revealed that students were less concerned with the quality of the peer assessment than the attitude of their peers giving the assessment. Young learners may not be a reliable source for assessing peers' work. Butler and Zeng (2014) found that peer-

assessment had limited application for younger students (9-10 years old), comparing to older students (11-12 years old). Students themselves are also concerned about their competence in peer-assessment (Harris & Brown, 2013). Although most of the students in this study were 11-year-olds, they still questioned the quality of peer-assessment. LGL, for example, did not believe in peer-assessment because of the bias among classmates. RT also emphasised that students' knowledge might be unreliable and may impact negatively on peers. However, Afitska (2014a; 2014b) and Pinter (2007) claimed that 10-year-old students are mature enough to assess their peers and themselves in a way that can promote their learning. It was not only students, but also teachers who doubted students' ability to assess their peers. As RT claimed, students were easily over-confident in what they thought they knew, whether it was right or wrong. It was also possible that students were not knowledgeable enough to make critical judgements. If cognitive burden on young learners is beyond their abilities, peer-assessment is risky (Butler & Lee, 2010; Butler & Zeng, 2014). Teachers could cut in and guide students. LT had relatively more confidence in students saying that they can make good judgements about the knowledge that they learned. Peer-assessment and self-assessment can only be effective when the teacher has entrusted the students to use the assessment (Harris, 2007).

Lacking guidance was the case not only in this study. More training and support were needed for students to turn their spontaneous judgements into effective peer-assessment and self-assessment. Other studies in China have pointed out that students' engagement and perceptions of peer-assessment tend to be low without teachers' careful planning, thoughtful training of students, and sustained support throughout the activity (Wang, 2014; Zhao, 2014; Chen, May, Klenowski, & Kettle, 2014). Studies have indicated that children younger than seven years old are excessively optimistic and naïve about their ability (Butler & Lee, 2010).

RT believed that children are innocent and naïve, hence they may give peer assessments candidly without any tact or euphemism. But it may cause another problem which is that some students with low self-esteem may dislike critical peer-assessment. What teachers can do is guide the students out of the awkward situation, while at the same time teach all the students how to deal with criticisms, and how to see people equally when assessing others. Learners do not welcome peer-assessment when they see it as criticism rather than help (Morris & Tarone, 2003). RT especially emphasised that the purpose of education was not just to teach the knowledge, but also to cultivate a positive approach to life. Teachers needed to and wanted to create a friendly environment for children to express themselves freely and equally. As RT suggested, to create a friendly and equal environment in the class is the priority. Students are more ready to express their opinions and risk making mistakes in a safe classroom environment (Jiang, 2014). Moreover, the ability to handle assessment critically is the key in peer-assessment, and this needs more training in students.

The teachers in the study have stressed the importance of equality among students as the precondition for healthy peer-assessment. The quantitative and qualitative student data revealed that peers' attitudes were crucial in peer-assessment, and teachers shared this view. RT has embedded her belief in every aspect of teaching: educate a person to become a free being who can maintain himself/herself and become a member of the society. This is

what Immanuel Kant called: practical education or moral education (Buchner, 1904). RT placed 'being a person who respects each other' ahead of 'being purely a mistake detective'.

6.2.5 Self-assessment

Self-confidence and strong willingness to improve probably made self-assessment the most useful FA strategy for students, particularly in HA and LA groups. According to the quantitative findings, self-assessment was the most influential FA strategy for all the students, especially in the HA group (Appendix, Table 4). HA students, who had greater self-confidence and better academic performance, were ready to be critical of themselves via their own assessment, but mainly they wanted to confirm their achievement; while LA students were eager to improve themselves, sometimes through negative self-assessment. However, MA students' tendencies were more likely to be in the middle of the other two groups' patterns. On one hand, they were not confident enough to confirm their achievement; on the other hand, they were not as keen as LA students to improve themselves. Besides, they worried about the criteria in assessing themselves. Hence, the impact of self-assessment on MA's motivation was smallest among the three achievement groups.

Students used self-assessment because they could build their confidence and were able to adjust learning aims. Although in other research it was found that higher achievers underestimated the quality of their work in self-assessment, while lower achievers overestimated its worth (Sung, Chang, Chang, & Yu, 2010), the findings were different in this study. Notwithstanding the doubt in self-assessment, all but one of the interviewees still felt motivated by self-assessment. HGL, MGL HGR all felt that they could gain confidence via self-assessment. Research has shown that when students analyse their own learning, their self-efficacy will be improved (Andrade, Wang, Du, & Akawi, 2009). HBR and MGR believed that self-assessment can help them to spot weaknesses and then set new aims. Self-assessment can contribute powerfully to educational development if students examine and comment on their own work critically (Crooks, 2001).

Self-assessment was not only an FA strategy for use in the class, but also a learning habit that can help students in the long term. LT said that students treated self-assessment differently according to their learning habits and learning motivations. RT believed that self-assessment showed students' ability, and was not merely tokenism. If students want to reflect on their English performance, it is necessary for them to assess in the context, rather than in decontextualised manner (Butler, 2018). The 'ticking time' in the textbook in this study was a good example of decontextualised assessment, because some students did the exercise only because the teacher directed them. English learners in colleges were found not to engage in peer- or self-assessment actively (Chen, May, Klenowski, & Kettle, 2014). In this study, some students saw self-assessment only as an activity in the textbook, rather than a spontaneous habit in other contexts. Awareness of self-assessment is what the students need more than prescribed self-assessment activities that students do not like.

Self-assessment was closely related to other FA strategies. MGR mentioned she needed learning aims when assessing herself. MGL said that she would use self-assessment after teachers' feedback. HBL would reflect on himself if he received negative peer-assessment.

HGR and HGL liked to compare self-assessment to peer-assessment, and then adjust learning aims. Other FA strategies, such as feedback, cannot work well unless students accept it and internalise it via self-assessment (Crooks, 2001). It not only assessed the knowledge, it also helped students to adjust themselves emotionally, especially after peer-assessment and teachers' feedback. However, self-assessment was used less than other strategies (Figure 4.12), especially in LA. The lesser use of this strategy may reduce students' interests (Figure 4.20). Considering the evidence that self-assessment has a strong positive impact on students' motivation, a goal for the future might be to increase the use of self-assessment.

Self-assessment was a way to understand oneself, but boys needed time to get to know and understand this method. As MGL said, "*there is no need to take other's judgement, peer-assessment brings me mental burden*", but self-assessment will bring her self-confidence. It was a process of "*truly know one self*". However, there exists an inevitable problem: the different standards among people. Students worried that tiny mistakes will be enlarged because there were no unified standards. Compared to the assessment from others, girls preferred self-assessment; boys valued self-assessment more than girls, but only in the post-test (Appendix, Table 6). There was a significant increase in the positive relationship between self-assessment and boys' motivation. However, boys did not show obvious affection for self-assessment in the interviews. Compared to girls who were open about their feelings on self-assessment, boys appeared bewildered because some of them said "*I don't know*" or could not explain why they were motivated. More boys showed more impulsivity than girls, but this trait may become weaker over time due to the gradual maturation of the nervous system (Côté, Tremblay, Nagin, Zoccolillo, & Vitaro, 2002). In other words, boys, compared to girls, would need more time to become more aware of themselves, as a sign of being mature. Hence, self-assessment may have greater impact on them when they are older.

The effect of the primary school children's self-assessment in academic achievement was not clearcut. Self-assessment had moderate correlation with achievement for FAL students (Appendix, Table 9) and girls overall (Appendix, Table 10). Students in FAL used more self-assessment at the end of the fieldwork, as did girls generally (Figure 4.15). The increasing use of self-assessment may be one of the reasons for the relationship. However, the impact of the self-assessment on students' achievement was small. In the interview, some students, for example FAL boys, believed that self-assessment can help them to spot the mistakes, so they are more willing to correct the mistakes identified by themselves. LA students, especially those from FAR, expected that self-assessment could help them with the next exam. On the other hand, many students were in doubt about the effect of self-assessment, because they were not confident enough to trust in their own self-assessment.

Students also expressed concern about the absence of criteria for self-assessment. Students did not report very positive attitudes towards self-assessment. They described their attitudes as "*comparatively like it*", "*just so-so*", or "*will not assess oneself after a bad exam*". Those who did not like self-assessment complained about the lack of assessing criteria. They literally "*do not know how to assess*" themselves, as MBL said. LGL also worried that "*sometimes you say you did a good job, but actually you did not*". Student assessment, including peer-assessment, seemed not reliable enough for students, because they tend to

believe the teacher's judgements. Dann (2002) suggested that students should understand the criteria and the process by which students arrive at their judgements. However, student assessment is not simply checking answers or making judgements, it is more about monitoring and evaluating the nature of their thinking (Cauley & McMillan, 2010). In this study, the schools did not provide enough information about student assessment for students to conduct it. Even if the students wanted to use peer-assessment or self-assessment, they could not find criteria to use as reference points.

Involving students in self-assessment can make FA productive (Black & Wiliam, 1998b), however, it was more useful and effective in terms of emotions and attitudes than in knowledge development. The positive effects of self-assessment on English performance was found in young learners in South Korea, but the effect sizes were rather small (Butler & Lee, 2010). First of all, primary school children may not be able to identify their own mistakes. An EFL learner who had been learning English for only one year, as in this case, was less likely to spot the mistakes without others' help. When the students use the language, they often use it based on the textbook or teachers' instruction and make mistakes unconsciously. If their peers or their teachers do not point out the mistakes, students hardly realise that they have made mistakes. Secondly, self-assessment's function was not just to examine the mistakes (Cauley & McMillan, 2010). Like peer-assessment, self-assessment often occurs spontaneously, such as in reflection. The format of it was not limited to the 'ticking time' in the textbook. It was important to develop the habit of introspection. Before the age of seven, children are excessively optimistic and naïve about their own ability; between eight and 12 years of age, their assessing ability improves, as does their self-reflection awareness (Butler & Lee, 2010). The participants in the current study were about 11 years old, at the stage of starting to realise their imperfections. It was easy to fall into self-doubt and become unconfident. However, self-assessment was a personal development method which does not only reflect on academic achievement.

Self-assessment was important to students as well as teachers. It was important to students because they can think critically in self-assessment, and review the knowledge learned previously. However, self-assessment occurred infrequently during the fieldwork and according to LT, 'ticking time' was relatively new to them. Other research also found that self-assessment is often new to students (Afitska, 2014a). Students need time and support to learn how to assess themselves, so do teachers (Wylie & Lyon, 2015). RT echoed this: teachers need to self-assess too when learning how to guide students in self-assessment.

As can be seen in the discussion, although the quantitative data did not show a strong correlation between FA and achievement, FA strategies were statistically significantly related to achievement. The strength of the relationship was probably affected by the length of the fieldwork. Making progress needs time. Although the present study did not show that FA has a strong impact on students' achievement, given the evidence of a robust relationship between motivation and achievement, the prospects of a stronger impact of FA on achievement over a longer time frame is very promising.

Considerations of what the FA environment can bring to the students refers to not only the environment that the teacher created, but also the assessment habits that students developed in the environment. However, developing habits needs time. In this study, there

were no noticeable differences between FAR and FAL in student assessment and motivation correlations. That is not to deny that if there is an FA-rich environment, the positive impact of FA may increase, even in such a short fieldwork, because the students have been immersed in the environment for at least one year.

Both teachers had already accepted the concept of FA, but in practice they implemented FA differently. There are similarities to a study in Cyprus where primary school teachers embraced the idea of FA, but used strategies mechanically without active engagement of their students (Antoniou & James, 2014). In the current study, RT explained her use of FA via constructivist education theories. As Heitink *et al.* (2016) suggested, teachers need to take a more constructivist view of learning. It was obvious that RT had more pedagogical knowledge than LT, who was still in the first phase of being a teacher, when they are improving their teaching, absorbing teaching beliefs and knowledge like a sponge. On the other hand, RT was in the second phase, when there is a levelling off in the teacher's acquisition of teaching skills (Hanushek, Kain, O'Brien, & Rivkin, 2005). Due to different life experience and teaching experience, the two teachers had different beliefs. Beliefs and attitudes are important for 'deep' implementation of FA (Heitink, Van der Kleij, Veldkamp, Schildkamp, & Kippers, 2016)

6.3 Motivation and achievement

Achievement was closely related to motivation theories. As mentioned in the section on Achievement Theory in Chapter 2, achievement was anchored in almost every motivation theory that applied in this study. In the following section, the relationship between motivation and achievement will be discussed via the comparisons across groups.

6.3.1 Achievement groups

Achievement was motivational only if students' academic achievement was not already high, as evidenced by the lack of a relationship between HA students' achievement and motivation during the fieldwork (Appendix, Table 12). This was probably because HA students are already satisfied with their achievement compared to other groups. The need for greater achievement, featured in Dörnyei's framework (Table 2.9), was not as strong as for other students. The outcome of their participation in the learning would not be as attractive to them, because within the class, they had already achieved a higher position in terms of academic achievement. As HA students, there was little challenge for them. That lowered their anticipated value of outcomes and incentive value of success (Table 2.10 L2 Motivation Framework by Williams & Burden (1997) and Achievement Theory in Figure 2.7). Therefore, improved achievement cannot motivate them anymore. However, it was not easy to sustain their high achievement if they were not motivated at all. Ten weeks may be too short a time for them to realise a dramatic change in their achievement. If the fieldwork could run longer, the HA group may care more about their achievement in the post-test due to the potential fall in their achievement that relative to other students, who are improving.

In contrast, LA students, who may not be happy with their achievement, were strongly motivated by it, but only in the pre-test. When students realised that their achievement was behind others, it was natural for them to wish to outperform their peers (Performance goal

in Goal Orientation Theory), to show their competence (Attribution Theory Table 2.5), and to improve the areas of language learning that they had not done well at. These reasons reflected the motivation factors: *Learning Environment*, *Learner*, and *Language* respectively. However, the strong willingness to learn the language did not last long. Progress needs time, whereas in this study, the 10-week fieldwork was not long enough to produce noticeable differences in achievement. Even if the LA students were motivated and then made efforts in learning English in the early stages, their motivation could not be maintained when the actual achievement did not fulfil their expectations. Their results belonged to the Individual: Individual/Action-Input motivation element, while their efforts are the output. From the students' perspective, they are easily demotivated when the output is greater than the input. This may explain why LA students' motivation was not related to their achievement in the post-test.

As for MA students, they were not as confident as HA students in their achievement, or as eager as LA students to prove themselves. Hence, their achievement and motivation were related to each other at both test stages, but only with a weak strength.

6.3.2 FA environment

Although the relationship was weak, the FAR students' motivation and achievement were significantly correlated throughout the fieldwork. FAL students had stronger correlations than FAR, especially in the pre-test, where their achievement was strongly related to *Language* and *Learner* (Appendix, Table 13).

A changing environment can refresh students' motivation to achieve. The fieldwork started at the beginning of a school term, when students came back from holidays. Their motivation in learning was higher than usual because the new beginnings brought them a fresh attitude. This may explain the stronger correlations in the pre-tests. Later in the post-test, children were familiar with the schooling pattern after ten weeks. This was why RT emphasised the importance of sharing learning aims. It was easy for students to feel 'lost' in the eight-lesson school day, not to mention a longer school term. Children will be motivated if they can receive new stimulation (Achievement Theory and Self-determination Theory in Figure 2.7). Learning aims, as a refreshing change in the dull study environment, would be a highlight for students to clear their minds for study. Other changes, including breaks, changing vocal tones, exaggerated gestures, *etc.*, all may serve to refresh the student and consequently stimulate their achievement.

A learning environment with plenty of FA can help to maintain the close relationship between motivation and achievement. In FAR, children were immersed in FA strategies. Compared to FAL, FAR can better maintain the relationship between students' motivation and achievement. Although the difference of motivation-achievement relationships between FAL and FAR was not significant, it had shown the trend of the two developing ways in the correlations. Of course, the environment was not the only reason why achievement and motivation were closely related, however, it may contribute to the maintenance.

6.3.3 Gender

Achievement and motivation correlations in boys were stronger than in girls in the pre-test, but weaker in the post-test (Appendix, Table 14). Moreover, most of the correlations were moderate. In other words, boys' achievement was closely related to their motivation, but then the correlation dropped down; meanwhile, girls showed the opposite trend in their motivation-achievement relationship.

Achievement has been proven as a variable that influences students' motivation. First of all, girls were proven consistently to outperform boys in standardised exams (Foster, 2000), especially in language (Wong, Lam, & Ho, 2002; King, 2016). Their achievement levels are higher than boys', evidenced by this study too (Figure 4.3). Secondly, in the context of China, due to the historical gender bias in China, boys were expected to achieve more than girls. Although some researchers have indicated that the one-child policy is beneficial to gender equality, for some observers this conclusion is too optimistic to be accepted (Liu, 2006). However, in studies, boys have shown lower levels of motivation, engagement, and achievement due to negative peer attitudes toward school (King, 2016). Therefore, it was possible that the conflicts between the expectation and the reality made the boys lose their interest in achievement soon after the beginning of the fieldwork. They were more likely to conform with the gender stereotype. Meanwhile, the better achievement helped girls' motivation to grow stronger.

6.4 Lower-achieving girls

In the analysis, it was obvious that there were some sub-groups that have significantly stronger relationships between FA and motivation, which deserved consideration under separate sections. The data in **Error! Reference source not found.**were extracted from Table 7 in the Appendix. LA girls were mainly interested in assessments that were led by teachers; LA boys were interested in almost all the FA strategies except corrective feedback; HA girls were mainly interested in a few teacher-led FA strategies, positive peer-assessment, and self-assessment; HA boys were interested in student-led assessments.

Table 6.1 Strong correlation between FA and motivation in sub-groups

	LA Girl	LA Boy	HA Girl	HA Boy
SLA	**.928**/.745***	**.566***/.460*	-/-	-/-
Q-O	**.664*/.744***	**.555***/-	-/-	-/-
Q-C	-/**.843****	.473*/.479*	**.686***/-	-/-
F-CF	-/-	-/.443*	**.539***/-	-/-
F-S	**.727*/.677***	**.599***/.467*	-/-	-/-
F-CR	**.628***/-	-/-	-/.499*	-/-
PA-P	**.724*/.738***	**.661*/.649****	-/**.846****	**.687***/-
PA-N	-/-	-.475*/**.502***	-/-	-/**.636***
sa	-/-	-/**.754****	**.635*/.851****	**.765*/.756***

** Correlation is significant at the .001 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

a/b = pre-test Pearson Correlation/post-test Pearson Correlation

- = not statistically significant

FA had the most significant impact on LA girls; this was probably because they were more sensitive than boys, more eager to prove themselves to others and gain the recognition from teachers and peers. They would like to seize every opportunity to improve themselves. They have clear goals to achieve, therefore strategies like sharing learning aims were especially useful to them. They are sensitive; hence, negative peer-assessment cannot function very well for girls, including HA girls. Meanwhile, the lack of confidence in the LA girls may contribute to reluctance and inaction in self-assessment.

Boys with lower achievement wanted to prove themselves too. However, they were not as sensitive as girls. It was possible that they were attracted by many FA strategies at first, but over the course of study, they gradually found that the strategies that suited them were student-led assessment. Compared to the girls in LA, the influences from peers and themselves were greater. This may be related to the finding that peers were a more influential factor in learning than teacher-led strategies.

Girls with higher achievement did not show the extremely strong willingness to outperform others as they were already at the top of the class. Hence, the FA strategies have less impact on them in general. These girls were highly affected by assessment from peers and themselves. It showed that HA girls care less about the teacher assessment, but more about student assessment. Meanwhile, they still rejected negative peer-assessment.

For boys with higher achievement, the only strong correlations were between their motivation and student assessment. Teacher assessment did not affect their motivation. This may be due to their high sense of self-efficacy and background of strong academic achievement. Being relatively insensitive to criticism may have made boys more tolerant of negative peer-assessment. They were more likely to accept critical peer-assessment when they experienced this kind of assessment frequently.

Low achieving girls are more easily to receive peer rejection and discrimination due to their achievement (Bakker, Denessen, Bosman, Krijger, & Bouts, 2007). On the other hand, boys' personalities were more likely to be straightforward and insensitive, their motivation was less affected by FA. They were prone to the influences from peers and themselves mainly, including negative peer-assessment that girls disliked. Compared to girls, boys were more willing to draw peers' attention (Takagishi, et al., 2015). However, the gender stereotypes cannot be taken for granted or remain unchallenged. Individuals cannot be ascribed particular personality characteristics due to their gender. Therefore, the findings and discussion about gender characteristics observed in this study are just conjecture, and the discussions have to be mindful of stereotypes.

To sum up, the higher the achievement, the greater the impact of student assessment on students' motivation; the lower the achievement, the greater the impact of teacher assessment on their motivation. It was probably because lower achievers were keen on improving themselves, to prove their abilities, but at the same time were not confident enough in themselves. Therefore, the teacher was the stronger and more reliable source of motivation. In contrast, because higher achievers may receive more recognition from

teachers, teacher assessment may no longer be influential. Peers and themselves therefore became more important sources of motivation.

In the current study, only girls with lower achievement had strong correlations between their achievement and motivation across all the sub-groups, especially in the pre-test. The data in Table 6.2 were extracted from Table 15 in the Appendix. LA girls' motivation and achievement were significantly and strongly correlated; while other groups had little relationship between their motivation and achievement, except for HA girls' motivation factor: Learning.

Table 6.2 Strong correlation between motivation and achievement in sub-groups

	LA Girl	LA Boy	MA Girl	MA Boy	HA Girl	HA Boy
Language	**. 700*/-**	-/-	-/-	-/-	-/-	-/-
Learner	**.755* /-**	-/-	-/.340*	.344*/.253*	-/-	-/-
Learning Environment	-/**.605***	.452*/-	-/-	-/-	-/**.501***	-/-
Motivation	.**712***/-	-/-	-/-	.269*/-	-/-	-/-

** Correlation is significant at the .001 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

a/b = pre-test Pearson Correlation/post-test Pearson Correlation

- = not statistically significant

The only information about lower-achieving girls I found in the literature was in Bakker's work (Bakker, Denessen, Bosman, Krijger, & Bouts, 2007), which found that peer rejection and discrimination are associated with lower achievement, especially among girls. That is to say, girls with lower achievement were a group of students who were sensitive to peer-assessment, especially negative peer-assessment. The learning environment may bring them more influences, as Table 6.2 indicates: only *Learning Environment* had strong correlation with their achievement in the post-test. For this special group, their motivation and achievement, as well as the use of FA strategies on them, were closely related to each other. This highlights one of the most significant findings and conclusions in this study - that FA strategies had the greatest impact on motivation and achievement in girls with lower achievement. It added the missing puzzle of the studies about lower-achieving girls using other FA strategies apart from peer-assessment to the literature.

6.5 The Impact on motivation and achievement beyond FA strategies

Many unexpected but valuable information about the impact of FA on motivation and achievement emerged during the interview. Two pieces of information were outstanding because they were either evocative or innovative.

6.5.1 The influence of mothers

MBR mentioned his mother in self-assessment. It was his mother who made strict rules in learning language, hence he used self-assessment based on his mother's rules. MGL also mentioned the help from her mother many times, from the quizzes at home to the tutoring her mother gave her. The assessment input from people around the students was very

important too. Home environment factors, such as parents' occupations, education, and learning facilities at home can affect students' motivation and achievement (Muola, 2010). In China, where there is a traditional viewpoint that "Men outside the home, women inside", children's education, being an 'inside issue', is placed on mothers' shoulders. Introducing FA into a SA environment could be problematic from parents' perspectives, as they may worry that FA is ineffective in improving students' scores, particularly in high-stake exams (Poole, 2016). However, the current research is about classroom-based formative assessment. The factors outside of the classroom were not within the scope of this study but they could be investigated in the future.

6.5.2 Games and assessment

Researchers might also turn their attention to the assessment via the games in the language class, and how formative assessment can be applied in language games. Two of the students mentioned playing games. Thanks to the development of technology, game-based learning has been the subject of a great deal of attention and research. However, the positive impact of game-based learning has been debated vigorously with studies that support or counter it (Qian & Clark, 2016). Students in primary education expect to experience curiosity, challenge, novelty and fun in game-based learning (Hainey, Connolly, Boyle, Wilson, & Razak, 2016). It is possible to stimulate learning if games are used as an assessment tool in the class. Unfortunately, this topic falls outside the focus of this study but would be a fruitful direction for future work.

6.6 Particular points for consideration in relation to the findings

6.6.1 The participants

Generally speaking, girls were more well 'behaved' than boys in the interviews. It was not just their answers, but also the way they answered the questions. Girls took every question seriously, but some boys were playful or hesitant during the interviews. But the biggest impressions were about the lower achievers. Compared to the more talkative MA or HA students, LA students were commonly shy and extremely careful in their responses. When they answered the questions, their voice volume was low, and they would sometimes change their answers if they noticed any expression on my face. I had to be extreme careful to avoid giving them any leading information.

In the questions on whether the teacher used a certain FA strategy, many students changed the answers. For example, HBR changed his answer from confirmative feedback to corrective feedback; MBR changed from confirmative feedback to suggestive feedback; MGR changed from corrective feedback to confirmative feedback; LGR and LGL changed from confirmative feedback to suggestive feedback. There were many reasons for the changes, for example, they had never thought about the question before and it was hard to find an answer at that moment; they did not understand the meaning of the terms, even after I explained it; they only wanted to answer the questions with 'correct' answers which can meet others' needs or expectations, *etc.* It was hard to tell whether the children were telling the 'truth', and this issue reflects the points covered in 'Doing research with children' in Chapter 3, that children are a special group of people. Therefore, in this study, the

discussion was premised on the belief that children were telling the truth if that was consistent with my observations during the fieldwork. If not, the possibilities for the discrepancy were discussed.

Many of the students' viewpoints were similar to what teachers thought. RT and LT had their unique perceptions of FA as teachers. In general, the relationship between teacher and students would influence the learning. Many students would come to the office after class to talk to RT due to her close relationship with her students. On the other hand, LT was less zealous than RT. During the fieldwork, RT and LT were very dedicated to their teaching, but with different teaching styles. RT would offer different rewards and opportunities for students. For example, "*as long as you complete your task, I will offer the chance to them to be a 'little teacher' in my class as they wish*". At the same time, LT was anxious about classroom management. Collaborative learning is often difficult to organise and implement effectively (Wylie & Lyon, 2015). The skills for effective management of classroom discipline therefore concerned LT.

6.6.2 Chinese culture in using and receiving FA.

There was an additional finding about teacher authority in China. When LGL talked about the aims, she said students can know what "*should*" or "*should not*" be learned via SLA. It seemed that she strictly followed the teachers' instruction and requirements, without her own opinions. This behaviour may be due to her personality, or lower self-esteem related to lower achievement, or the teacher's authoritarian teaching style. Regardless of the reason, this phenomenon may evidence the rigid teaching and learning environment and hierarchical relationship between teacher and students in China.

Mianzi (face) and Guanxi (relationship) are serious issues when discussing the interactions between Chinese people. Higher educational performance has become a significant marker for Mianzi, which is influenced by the social pressures the only-child families have in passing the Mianzi to the next generation (Guan & Ploner, 2018). Because the potential failure in academic achievement could let them lose face, hence, students may try to save Mianzi and keep a good Guanxi for oneself and each other in FA, peer-assessment and self-assessment in particular. The embarrassment caused by making mistakes in front of classmates was the kind of anxiety seen in language classrooms all over the world (Gardner, Tremblay, & Masgoret, 1997). Although the teacher mentioned that young people are naïve and straightforward, in the interviews, it was not difficult to realise that students may avoid the embarrassment of losing Mianzi and keep the harmonious Guanxi with classmates and teachers by holding back true feelings. As the HGL said "*pupils are too straightforward, then what they said may hurt one's pride*", and the MBR admitted "*it [peer-assessment] feels so shameful*". In this study, the Chinese culture may contribute to the explanation of certain relationships and behaviour, such as the avoidance in negative peer-assessment.

6.6.3 Nuances in translation

It was worth noting that there were nuances in Chinese data and English translation. The students and teachers were using Chinese to answer the questions in interviews and questionnaires. After analysis, the data were then translated by machine and the author

into English. However, translation is not word to word exchange, because there seldom is a one-to-one relationship between languages, especially when it comes to cultural nuances. For example, the phrase 'negative peer-assessment' in this thesis referred to several circumstances which included critical peer-assessment, criticism, or even discrimination. The tone of voice can sometimes mislead as well. For instance, it was difficult to tell if an interviewee was joking or being sarcastic from a written transcript. The transcript was translated by machine translator such as Google translate, and adjusted by subjective human translator. Nevertheless, without other opinions on the translation, it may cause possible distance between what the interviewee wanted to say and what the thesis reader understood.

6.7 Summary

This chapter discussed the findings in Chapter 4 and Chapter 5, following the main themes of FA strategies, achievement and motivation. The unexpected findings, surprising results and a few additional considerations were discussed as well.

The results of this study show that FA has positive and significant impact on students' motivation and achievement. Teachers and students showed positive attitudes towards formative assessment as a whole, but were critical about each strategy.

After combining the quantitative and qualitative evidence together, and linking them with earlier literature, it is clear that FA, a concept adopted from the UK and a successful assessment practice endorsed by many western countries, is an adaptable and useful assessment means which can apply in English lessons in Jiangsu Province in the exam-oriented Chinese context. Teachers can use this practice more effectively by targeting different group of students with varied strategies. To be specific, formative assessment is the most effective for girls with lower achievement. Among all the FA strategies, student assessments, namely peer-assessment and self-assessment, are the most popular among students.

In the next chapter, a conclusion will be made to synthesise the whole thesis from literature to methodology and findings.

Chapter 7 Conclusion

7.1 Why complete this study

This study started 'from scratch' out of the researcher's personal interests and was subsequently developed and continued in light of the research gap in the literature. However, it was during the fieldwork that the actual need for formative assessment in classroom practice became even more evident and gave the impetus to push this work to its final phase.

Before this study is published, there are at least two teachers who have a much better understanding of FA and how to use it. Working with the participating teachers for a period of 10 weeks enabled both the researcher herself and the teachers to reflect carefully on the nature of formative assessment. The findings in this study will be reported back to the teachers and also to some of the student interviewees who requested the final reports. The research process has provided an opportunity for the teachers to learn more about the impact of formative assessment and to adjust their teaching accordingly. It is worth noting that all the FA strategies studied in the current research took place in the usual schedule of daily teaching, with little extra time commitment required. Teachers are more ready and willing to use FA in these circumstances. As mentioned in 3.5, all the English teachers in the school gather together weekly to prepare the lessons collaboratively. It is therefore also likely that other teachers may benefit directly from the enhanced FA knowledge of the two participating teachers if they share their experiences and advice.

Of course, the influence is limited if it only depends on these two teachers or indeed all of their colleagues in their school. More people will benefit from the findings if this work is published online and is open to other researchers and teachers. Further possibilities for dissemination of recommendations based on the evidence in this study would include the in-school teacher CPD (Continuing Professional Development) training about FA and also workshops in class for students to train them how to do effective peer-assessment and self-assessment (Jones, 2014).

Moreover, this research is rooted in the primary education setting, so it particularly enriches the corpus of FA literature that has so far been dominated by the higher education context. Finally, this research has provided a case study for the 'Anglosphere' about the impact of FA in the context of China, which can be used for cross-cultural case comparisons.

7.2 How this study was conducted

When I started this project, my interest in formative assessment was strong. Because of my previous work on FA in Master's degree study, I already had a good start-up knowledge base for this project. From that point, the literature on motivation was read and reviewed comprehensively, in order to build my own motivational framework tailored for this study. After getting in touch with the potential participating schools and teachers in China, the ethics application was submitted with the draft research plan to the University of Sheffield Research Ethics Committee. Unfortunately, the Committee rejected the first ethics

application due to a misunderstanding deriving from cultural differences regarding the issue of access to the schools. There was a key concern about how the research and the researcher would be understood in China, especially since the work would be undertaken under the authority of a university outside of China. As motioned in 3.7.1, *Guanxi* is the principle underlying the method to gain support from the Chinese schools and teachers in such an individual-led research project. Therefore, the ethics application was revised and re-submitted, but in the meantime, the fieldwork schedule had to be rearranged.

After gaining the approval from the University's Research Ethics Committee (Appendix 7), the fieldwork commenced, running from February to May in 2015. During this three-month period, there were a one-week pilot study and a ten-week main study. After the pilot study, research tools were adjusted according to teachers' and students' feedback and then applied in the main study. During the fieldwork, I attended one English lesson in the participating classrooms per day, i.e. four lessons per week. Notes recorded during observations were not used as primary data in this study, but were one of the sources of information used to cross-reference and validate the data that teachers and students provided in the interviews. The primary data collected during the main study were: a) the quantitative data on the use of FA, the attitude towards FA, the motivation in English learning, and achievement scores over ten weeks for 196 students; and b) qualitative data gathered from interviews with two teachers and twelve students about their perspectives on FA.

In the data analysis phase, the quantitative data were entered and calculated in SPSS. Motivation was analysed and categorised as three factors: *Language, Learner* and *Learning Environment*. The main findings focused on the correlations between FA strategies, motivational factors and achievement. The qualitative data were transcribed from audio into text first, then coded and themed in NVivo (see an example of coded transcript in Appendix 12.1). The main findings focused on teachers' and students' opinions and attitudes towards each FA strategy and their ideas on other related matters.

In the writing up phase, the first task completed was the methodology chapter, because the fieldwork experience and its procedures were relatively fresh in the memory. The second completed task was the literature review which was updated on the basis of the confirmation review process two years prior. The revised literature review incorporated a large amount of new or recent evidence especially in the areas of formative assessment and achievement. The material for Data Analysis and Findings was divided into two chapters, quantitative and qualitative, for organisational clarity in reporting the two rich sets of data. The structure of the chapters was based on the five research questions in 3.2, and used again in Chapter 6 Discussion. The last sections completed in the writing up process were the Conclusion and the Introduction. There is a Chinese saying about what makes a good article: the beginning should be as pretty as phoenix's head, the main body should be as rich as a pig's belly, and the ending should be as powerful as a leopard's tail. Therefore, the introduction and the conclusion were kept brief and to the point.

Writing up a thesis is not a one-off piece of work or a linear process. During the writing phase, more literature was added, unnecessary sections were taken out, and each chapter was edited and revised thanks to the help of my supervisor. New literature in all the sub-

fields and topics was developed and published over the course of my doctoral study, therefore the process of updating the literature review not only allowed a deeper understanding of the themes and the gaps I identified (or missed) before the fieldwork, but also enabled me to capture the very latest relevant research, such as some recent studies in China (OECD, 2016; Poole, 2016; Peng, 2017; Liu & Xu, 2017). It was also the experience of talking to many English teachers in China, including the two teachers who participated in the fieldwork and those I met at conferences, that made me rethink the scope and direction of the literature review, as well as the emphases in analysis. Therefore, some sections about FA frameworks were deleted, because they were judged to be not relevant to the analysis in the later chapter.

In all these ways, the work of the past four years evolved into a complete thesis.

7.3 How to address the research questions

This study investigated the three-way relationship between FA, motivation and achievement in language learning, more precisely in the context of primary schools in China. It used triangulated sources of data to provide an empirical case to support the positive impact of FA on students, within the wider debate on FA. The case study fills the gap in the body of empirical studies of FA in primary education, in the Chinese context but in the Anglophone literature, and particularly with regard to implementation of concrete FA strategies. The findings suggested that FA can be applied in language classes in China, and also bring benefits to students' motivation and achievement. The impact of each FA strategy varied in different groups of students. These variations led to evidence-based suggestions to teachers and relevant practitioners about how to use FA more effectively and in more targeted ways.

Is there a relationship between FA strategies and students' motivation?

First of all, it was proven that there was a strong correlation between FA and motivation in this study, especially in lower achievers (LA) and higher achievers (HA). Compared to FA strategies related to teachers, those related to students were more useful in motivating students (4.6.1).

In terms of achievement level, LA show stronger correlation between FA and motivation than HA in the pre-test, but weaker in the post-test (summary 1 in 4.6.2, p.100). FA had a greater impact on LA motivation in the pre-test, but on HA in the post-test (summary 2 in 4.6.2, p.100). The number of effective FA strategies was greater for LA than for HA (Appendix, Table 4). Basically, every FA strategy was useful to LA. But it was only self-assessment and positive peer-assessment that had strong correlations with HA motivation (summary 3 in 4.6.2, p.100). With reference to the motivation factors, *Language* and *Learner* were important to LA all the time. However, they were important to HA only in the beginning. HA students were more influenced by FA in terms of *Language* and *Learning Environment* (summary 4 in 4.6.2, p.100). As a group, MA students did not show any strong correlation between their FA use and motivation, so the FA used in the classroom does not appear to serve medium achievers well. However, these correlations in the MA group, although only moderate or weak, did show increasing strength over the fieldwork (Appendix, Table 4).

In terms of FA environment, correlation between FA and motivation was stronger in the FA-rich environment (summary 1 in 4.6.3, p.101). All the FA strategies had positive impact on students' motivation; however, only a few produced strong correlations (Appendix, Table 5). In FAR, motivation had strong relationships with positive peer-assessment, self-assessment, and suggestive feedback; in FAL, motivation was influenced mainly by self-assessment. Student assessment, and self-assessment in particular, were the most effective FA strategies that increased students' motivation (summary 3 in 4.6.3, p.102). When it came to motivation factors, students in both FAR and FAL felt motivated in the light of *Language* and *Learner*. But in the post-test, FAR students' motivation shifted slightly to *Learner Environment*. This may suggest the positive impact of the use of FA in FA-rich environment (summary 2, 4 in 4.6.3, p.102).

In terms of gender, FA-motivation correlation did not show significant differences between girls and boys (summary 1 in 4.6.4, p.103). Boys in the pre-test showed strong correlations between FA strategies and *Language* and *Learner*, but the relationship was weakened in the post-test. Girls' FA-motivation relationship did not change very much (summary 2 in 4.6.4, p.103). Self-assessment was the most effective FA strategy for girls before and after the fieldwork, and was important for boys only in the post-test. Positive peer-assessment was important to girls too (summary 3 in 4.6.4, p.103). Concerning motivation factors, like other comparisons, *Language* and *Learner* showed more and stronger correlations with FA than *Learning Environment*. Boys were not motivated by FA in terms of *Learning Environment* in the pre-test, but it started to have more salience in the post-test (summary 4 in 4.6.4, p.103). During the fieldwork, boys' FA-motivation relationship changed a lot in many ways, while for girls it was relatively stable.

To extend the research findings, the correlation test was also run within the gender and achievement level matrix sub-groups. It emerged that female LA students' motivation was the most likely to be boosted by FA, followed by male LA, female HA, and male HA. These findings are summarised in Table 4.7.

Is there a relationship between FA strategies and students' achievement?

Secondly, it was proved that the relationship between FA and achievement was significant, but not strong (4.7.1). The impact of FA on achievement was either moderate or weak if compared within different achievement groups (4.7.2), different FA environments (4.7.3), or different genders (4.7.4). However, after running a correlation test within sub-groups between achievement level and gender, it was very clear that FA - and teacher-related FA strategies in particular - was strongly correlated with LA girls' achievement (4.7.5).

Is there a relationship between students' motivation and achievement?

Thirdly, it was proved that motivation and achievement were significantly and moderately related. The relationship in the pre-test was stronger than that in the post-test (4.8.1).

In terms of achievement level, the relationship between LA's achievement and motivation in the pre-test was at the border of moderate and strong. However, LA did not retain the

significant relationship at the post-test. MA group had significant but moderate correlations between *Language/Learner/*motivation and achievement. Their achievement was not related to *Learning Environment*. These MA relationships did not change much during the fieldwork. HA, on the contrary, had no relationship between their motivation and achievement at all (4.8.2).

In terms of FA environment, FAL students may recognise achievement as more important than FAR students do, because *Language* and *Learner* in the FA-lacking environment were strongly related to achievement in the pre-test. However, in the post-test, the relationships between motivation and achievement in both FAL and FAR turned weak. The strength of FAR's motivation-achievement relationship did not change very much over the fieldwork. Within all the motivation factors, *Learning Environment* had no relationship with achievement during the fieldwork in any FA environment (4.8.3).

In terms of gender, boys' motivation-achievement correlations were stronger than girls' in the pre-test, but weaker in the post-test. Boys' motivation-achievement correlations were decreasing, while for girls they were increasing. Most of the relationships were of moderate strength, except for boys' achievement, which was strongly related to the motivational factor *Learner* in the pre-test. However, in general, boys' motivation was not related to their achievement in the post-test (4.8.4).

Consistent with the earlier analysis, correlation tests were run within the sub-groups of gender and achievement level. Not surprisingly, the data showed that LA girls' achievement was strongly related to their motivation in the pre-test. But the correlation was not statistically significant in the post-test. In terms of motivation factors, the impact appeared to have changed from *Language* and *Learner* to *Learning Environment*. Therefore, LA girls emerged as the crucial group warranting further attention in studying the impact of FA on students' motivation and achievement (4.8.5).

What is the impact of FA from English students' perspective?

Generally speaking, students showed a positive attitude towards most of the FA strategies, apart from peer-assessment. They like sharing learning aims because the aims provide them with a sense of direction which can provoke their curiosity about the lesson and even the language (5.2.1). They like questioning and prefer open-ended questioning because they can express themselves freely in these questions and learn from the varied answers (5.2.2). They like feedback and prefer confirmative and suggestive feedback because confirmative feedback can bring them confidence and encouragement; meanwhile, suggestive feedback can inspire them and help them study actively (5.2.3). Higher achievers and medium achievers like positive peer-assessment because it is inspiring and encouraging, while lower achievers prefer negative peer-assessment because they want to prove themselves to others more after critical peer-assessment. Some of the students did not like peer-assessment, because they believed that different people apply different assessing standards, and they perceived this as unfair (5.2.4). They like self-assessment, from a knowledge perspective because they can find their mistakes and the learning leaves a deeper impression, and from the emotions perspective because they can comfort themselves or feel proud (5.2.5).

What is the impact of FA from English teachers' perspective?

Teachers provided a different perspective on each strategy. They like sharing learning aims because it can help students build up confidence and develop learning habits. Also, sharing learning aims can help the teachers to guide their own teaching (5.3.1). They like questioning, preferring open-ended questioning. They believe that close-ended questioning is the foundation before asking open-ended questions. In open-ended questions, students can reconstruct their knowledge, which can ignite their passion in learning (5.3.2). Teachers also like feedback but insist that feedback cannot be empty. The veteran teacher suggests that it is better to assess the behaviour than the person (5.3.3). They like peer-assessment, preferring negative peer-assessment because it is more effective. In conditions of equality among students, peer-assessment can help students not only in the language learning, but also in life lessons (5.3.4). Teachers also like self-assessment because students can learn to be critical when they are reflecting on themselves (5.3.5). Peer-assessment and self-assessment can teach students "learn to learn" (Black, 2015).

The answers to the quantitative research questions are all positive. Moreover, there is a statistically significant positive relationship between them. Except for female lower achievers, the relationship is not strong. It makes the lower achieving girls the group that can benefit the most from FA strategies.

Generally, students and teachers held a positive attitude towards FA strategies, but students have mixed feelings about corrective feedback and negative peer-assessment. The traditional Confucian ideals may contribute to this phenomenon, that it is important to keep good Guanxi with people around and give them Mianzi by not commenting negatively. Nevertheless, they still believe that FA is an effective method to motivate students and improve their achievement.

7.4 Original contribution to knowledge

Firstly, the quantitative data prove that each strategy has a different impact on different groups of students' motivation and achievement, therefore, there is a need for well-differentiated FA. It should be used wisely in different groups. Students have their preferences in relation to FA strategies; and they share responsibility with the teachers for successfully implementing FA. It is not only responsibilities, but also young students' willingness to be independent and have more control of their own learning. The data also indicate that the preference for FA and the impact of FA on achievement are not always positively related. Some strategies that students do not like, e.g. corrective feedback, can actually improve their achievement more than other types of feedback.

Secondly, teachers and student share some common opinions of FA. They all believe that learning aims are a lighthouse which can lead students in their learning. Without the aims, the students may feel lost in language learning, and the teachers may lose the track of their teaching. To both teachers and students in this study, non-verbal feedback is very important than has been recognised. In the context of long school hours in China, physical feedback,

such as facial expressions, eye contacts, hugs, high fives, etc., may stimulate more than oral feedback.

Sometimes, teachers and students see FA in a different way. Students dislike corrective feedback, meanwhile, the example that teachers using skilful corrective feedback indicates the need to use corrective feedback wisely, especially when use it as a recast positive feedback in effective language teaching. Students dislike close-ended questioning, however, teachers find it is important to ask memory questioning too. The dichotomy of open- and close-ended questioning is over-simplified when discussing the forms questioning might take. The focus on different functions in questioning may bring the teachers new ideas of using them accordingly, rather than using one particular type only. The conceptualising by young learners of FA is very insightful so, where teachers think in a straight-jacket of learning aims, the students see this as sharing ideas. They think peer-assessment can sometimes be fake and confirmative feedback could be excessive. The consistency of the FA concepts in both learners and teachers is particular important, as they may understand each other's' needs only if they are on the same page. To communicate, and to exchanges learning and teaching reflections are an effective way to bring teachers and learners together for a better learning experience.

Thirdly, teachers and students have their thoughts about FA beyond each strategy. Teachers believe that FA is not merely an assessment method, but a learning habit. It is part of the nuanced, relational, 'life lessons' that can be embedded in students' character-building. A lighthouse is the image of not only the learning aims, but more importantly the FA. The personality differs in each individual. Students learning with a warm heart and teachers teaching with love are the key affective points when implementing FA.

Lastly, in terms of methodology, there are contributions to doing research with young children and gaining their valuable insights. Children, as a special group of participants in academic studies, need more patience and time than other age groups. In order to gain their trust, having a good relationship with young children is vital in a study like this. It will take a longer time, yet the results are rewarding because children would talk more and talk frankly if they trust you. When collecting data in primary schools, it was obvious that learners were nervous at first. They were more comfortable talking to the researcher in the second week. The time spent with learners before the interviews was worth it. The language used with them needs to be age-appropriate, which means children need to understand the instruction or questions provided, rather faking their answers to appeal to the researcher(s). The researcher found that a few interviewees tried to be positive in some questions, yet their answers were not consistent in the follow-up questions. The questions, hence, were explained again in a different way, and were asked again. Researchers may need more patient to adjust the research tools when working with young learners.

7.5 Implications for practice

It is a researcher's duty to find ways to build interactions between the academic world and the world of practice. Otherwise, as commented by a teacher at the EFL Teaching and Assessment Conference (see 2.1.9) the gap between research and practice will never be closed. Throughout the work with teachers during the fieldwork in 2015, I have come to

understand with greater depth that FA implementation was much more than a simple process of applying each FA strategy to all students. There are many tips that need teachers' attention when implementing FA in primary schools in China.

Due to the nature of FA strategies, some strategies, such as peer-assessment and self-assessment, are used less than others (Lysaght & O'Leary, 2013; Afitska, 2014a; Wylie & Lyon, 2015). However, in the analysis and discussion above, it was noticeable that student assessments had their special role in teaching and learning, which cannot be underestimated.

7.5.1 Using FA strategies

Sharing learning aims

SLA can be used to motivate LA students, especially LA girls (Sharing learning aims in 6.2.1, p.135). More attention needs to be paid to boys' engagement with SLA since they showed a tendency to lose interest in SLA more easily (summary 2 in 4.6.4, p.103). A generally FA-rich environment is strongly recommended, because the FA-lacking environment may reduce or remove the positive impact of SLA on motivation (summary 3 in 4.6.3, p.102). However, unlike the picture for motivation, achievement was not much influenced by SLA in this study, except for the LA girls (Appendix, Table 11).

Setting appropriate levels of difficulty in learning aims needs to be a priority in teachers' planning and preparation (Sharing learning aims in 6.2.1, p.135). Students in different achievement groups respond to the shared aims according to the degree of difficulty and how interesting they find the aims. These issues influence whether the students are motivated to continue with the study. Such findings may be useful to textbook designers too, given that, to the researcher's knowledge, there are no clear hierarchical or stepped aims in the English textbooks currently used in China (Sharing learning aims in 6.3.1, p.**Error! Bookmark not defined.**). If there is a 'one size fits all' approach to learning aims, higher achievers may find them too easy but lower achievers may find them too difficult. Both scenarios may lead to amotivation in students. Therefore, it is vital for teachers to reflect classroom realities by differentiating the learning aims in the curriculum.

As RT and LT said, sharing learning aims can guide teachers' own teaching, which meets teachers' personal goals (Harrison, 2005). Teachers' sharing of learning aims not only gives directions to children, teachers can also reflect on their teaching at the same time. However, SLA is not merely giving out a note or telling students what to do at the start of the class. It is for sharing the learning purpose and developing quality success criteria.

When teachers share learning aims with students, students feel the 'ownership' of study, if the students are engaged in the goal setting (Curry, Mwavita, Holter, & Harris, 2016; Harris, 2007). Therefore, the strategy is not only for conveying the aims, but more importantly, to engage students in the process of setting learning aims.

For HA students, the aims need to be challenging, and for LA students, the aims need to be within their abilities. However, the LA students cannot be stuck in a spiral of low

achievement. All students need to use *stepped* learning aims to achieve something beyond their current level. It is essential to set exemplars, so the students know clearly what they are pursuing. Boys in particular may lose interest in SLA easily, so it is very important to ensure that the aims are interesting and engaging. Of course, being interesting or not also depends on the textbook. Therefore, the appeal of textbook content and organisation can also be considered when books are in development. Finally, if sharing learning aims is not already part of regular classroom practice, embedding the strategy is recommended. After all, students are readily motivated by SLA.

Questioning

Questioning can be applied to motivate students, particularly in the LA group. (Questioning in 6.2.1, p.137). More attention can be paid to the effective use of questions with LA girls and boys, given that LA girls tended to show less interest in questioning at the beginning, specifically close-ended questioning, while LA boys seemed to be less engaged by open-ended questioning after a period time of study (Appendix, Table 7). The use of open-ended questioning influenced students' achievement. Moreover, its impact was greater than that of any other FA strategy (4.7.1). HA boys' achievement was more sensitive to close-ended questioning; while LA girls' achievement was largely influenced by open-ended questioning (4.7.5).

The types and contents of questions are clearly important, but so is the waiting time after each question (Questioning in 6.2.5, p.**Error! Bookmark not defined.**). Although time in class is limited, especially precious given the larger class sizes in China, teachers still need to give enough time for students to think about their answers. Rushed judgements cannot bring positive impacts to students. Even so, time needs to be spent equally on every student (Harris, 2007), so the teacher must judge fair waiting times to best balance everyone's needs.

Questioning is only considered an FA strategy when it is aimed at diagnosing learning and has follow-up activities to use the information that is collected by questioning (Jiang, 2014). It is important to check the mastery of knowledge via questioning, but beyond the memorial facts, it is more important to trigger students' thinking. The answers to a question cannot stay as a right/wrong dichotomy only. A good open-ended question needs to probe for deep understanding, on the basis that a teacher knows the students well, and create good opportunities for students to link their previous knowledge to the new learning.

Waiting time cannot be the reason that lower achievers receive fewer questions. It is common that new teachers like LT in this study worry about the time 'wasted' on a few students who cannot answer questions quickly. However, it is not fair to lower achievers if they experience fewer challenging questions only because the teachers do not want to waste time in the class. A variety of questions could help students to answer. Teachers need to avoid rushed judgements (Torrance & Pryor, 2001). However, it is not fair for other students to wait for too long, so a balance needs to be struck. Other FA strategies, such as suggestive feedback and peer-assessment, would be useful methods to apply in this situation.

It is important to shift students' negative attitudes towards close-ended questioning and their utilitarian attitude towards open-ended questioning. Both teachers agreed that each means of questioning has its role, however, students believed that open-ended questioning is superior to close-ended. Moreover, the purpose of open-ended questioning is more than to "*prepare more answers for the exams*". Students need to learn to think in open-ended questioning. As well as paying attention to the questions themselves, teachers need to be mindful of students' sense of self-efficacy and help them to embrace challenges.

Feedback

Suggestive feedback needs to be used more, given that it was the most effective feedback strategy in triggering students' motivation, especially LA girls (Girls' LA group in 4.6.5, p.105). Corrective feedback had the smallest impact, and particularly on LA boys who did not seem affected by corrective feedback at all (Appendix, Table 7). LA students tended to lose their interest in corrective feedback over time, but feedback increased HA students' motivation over time (summary 2 in 4.6.2, p.100). Therefore, it is important to try to maintain LA students' motivation via feedback as a block of learning progresses. It is also important to ensure an FA-rich environment because this can increase the positive impact of feedback on students' motivation (Appendix, Table 5). HA boys showed that feedback had little impact on their motivation, so other strategies need to be considered for this group to motivate them better (Boys' HA group in 4.6.5, p.105).

Among all the types of feedback, only corrective feedback was effective in promoting students' achievement, but only in the LA group (4.7.2). The use of different types of feedback may affect students' achievement accordingly. In FAR, more confirmative feedback and suggestive feedback were used, and these were the feedback types that affected students' achievement. Similarly, in FAL where more corrective feedback and less confirmative feedback were used, it was corrective feedback that affected students' achievement, and the positive impact of confirmative feedback was smaller than that in FAR (Figure 4.13 and Appendix, Table 9). Confirmative feedback can improve girls' achievement, whereas suggestive and corrective feedback are more useful to boys in improving their achievement (4.7.4). Similar to SLA, open-ended questioning and three types of feedback were the most effective strategies for improving LA girls' achievement (4.7.5).

Finally, there is another function embedded in feedback, which is to stimulate or rouse students during a long school day. Therefore, body language and vocal intonation *etc.*, can be employed during feedback to engage students and keep them alert (5.3.3).

"The most powerful single moderator that enhances achievement is feedback" (Hattie, 1999; Hattie & Timperley, 2007). Feedback is an essential FA strategy that both teachers and students need to be careful with. First of all, teachers need to give confirmative feedback but cannot overuse it. Confirmative feedback is the most often used FA strategy in the class, partly because students like it. Sometimes it could be as simple as the word "good". However, the main feature of confirmative feedback is that it confirms students' progress compared to earlier performance (Crooks, 2001). Its impact is obvious and instant. It can also enhance students' knowledge by identifying the completion of tasks (Torrance & Pryor,

2001). The confirmation needs to be credible, rather than a simple comment given even when the students are not making an effort (Cameron, 2001).

Secondly, LA students need more suggestive feedback, although corrective feedback affects their achievement the most. In this study, LA is the group that received the least amount of suggestive feedback. However, they expressed strong willingness to improve their learning if they received suggestive feedback, probably because it identifies the extension of improvement and gives specific suggestion about how to improve (Crooks, 2001; Torrance & Pryor, 2001; Harris, 2007).

Thirdly, the prejudice against corrective feedback need to be adjusted. The impact of corrective feedback on students' achievement was significant, especially in the LA group. It is the teacher's responsibility to use it wisely. Here, 'wisely' refers to the teachers' attitude when using corrective feedback and the importance of using it only when and where it is appropriate. It is important for students to see corrective feedback as help rather than criticism. Feedback needs to target behaviours rather than the individuals. It needs to be clear that it is not a comparison with other students, but a comparison with the student's earlier work (Crooks, 2001). Young learners, especially girls, need positive feedback to recognise their achievement and build their self-confidence. Feedback is not about labelling, but should be delivered sincerely to help students to progress (Harris, 2007).

Fourthly, teachers can embed their personalities in their feedback, but this is not strictly necessary. Life lessons, exaggerated vocal tones, and physical gestures and actions are teaching habits that reflect strong personal characteristics. It is undeniable that they may have positive impact on students' learning, and on their motivation in particular. However, such habits are unique, and cannot necessarily be applied in another teacher's class.

Peer-assessment

Peer-assessment can motivate LA and HA students more than MA students (Appendix, Table 4). Generally, positive peer-assessment could be used more often than negative peer-assessment, in line with students' attitudes towards this strategy (Figure 4.16). However, the positive impact on motivation of negative peer-assessment increased hugely in the LA group over time, especially LA boys (Boys' LA group in 4.6.5, p.104). Meanwhile, care could be taken when using negative peer-assessment with girls as they showed they were not motivated by this strategy at all (Peer-assessment in 6.2.1), but appropriate negative peer-assessment can be used with boys to motivate them. As to improving achievement, positive peer-assessment was useful, especially to MA girls (4.7.5). On the contrary, negative peer-assessment did not affect students' achievement in any sub-group (Appendix, Table 7).

Proper training is needed for students to assess their peers and themselves (Peer-assessment in 6.3.1, p.**Error! Bookmark not defined.**). In peer-assessment, the attitudes of the assessor seem more important than the contents of the message. The quality or accuracy of peer-assessment has been questioned by students and teachers, therefore more attention is given to their feelings when receiving peer-assessment (Peer-assessment in 6.2.4 and 6.2.5, p.**Error! Bookmark not defined.** & p.**Error! Bookmark not defined.**).

When it comes to peer-assessment, attitude is the most important thing to consider. The ability and skill of students to peer-assess has been questioned by both teachers and students, so the function of peer-assessment can be seen as primarily emotional. For girls, who are more likely to be sensitive, and for students with lower achievement, 'how' is more important than 'what' when their classmate assesses them. The recognition from peers acts as an encouragement for the student to continue their study. That is the reason why most of the interviewees preferred positive peer-assessment. Teachers need to guide students, not only in terms of linguistic knowledge, but also moral development. However, not every teacher is able to embed the moral education in their teaching, since their priority is to teach English. This may improve when the teachers can free themselves from reliance on the textbook after more years of teaching experience. By then, they would have more flexibility in their approach to teaching.

The right attitude can make peer-assessment more acceptable, but explicit criteria makes it effective. Considering the size of classes in China, it is more efficient for teachers to organise teaching by using peer-assessment. Rather than asking students to applaud, teachers needs to teach students how to peer-assess properly. To do this, teachers need to use concrete examples, as well as provide support during the assessment (Black, 2015; Liu & Xu, 2017). Whether formally or informally, students need equal opportunities to give or receive the perceptions to or from their peers. Explicit structures, guidance and routines ensure the effectiveness of collaborative work (Wylie & Lyon, 2015). It may take longer for peer-assessment to have significant positive impact that is reflected in students' achievement. It is not only the assessment routine, but also learning habits that students can acquire via peer-assessment. Good learning habits would play a more important role in future learning, when more learner-centred study occurs.

Self-assessment

Self-assessment was more effective in motivating HA students than MA and LA students (summary 3 in 4.6.2, p.100). LA girls were not affected by self-assessment at all (Girls' LA group in 4.6.5, p.105). The impact of self-assessment was not affected by FA environment. Generally, the positive impact of self-assessment on motivation increased over time (Appendix, Table 5). Therefore, self-assessment was a useful FA strategy even in independent learning. This strong impact of self-assessment on students' motivation was apparent in girls from the beginning of the study, but boys showed the positive impact of self-assessment more gradually (summary 3 in 4.6.4, p.103).

In terms of student achievement, self-assessment had a positive influence only in the FA-lacking environment (summary 3 in 4.6.3, p.102). In such an environment, it may offset the lack of teacher-led FA to some extent. Self-assessment was not effective in improving achievement (Self-assessment in 6.2.2, p.**Error! Bookmark not defined.**). The only significant finding in the relationship between self-assessment and achievement was in the gender comparison: the relationship was slightly stronger in girls than boys (Appendix, Table 10).

Clear assessment criteria may enable students to use self-assessment appropriately (5.2.5). Due to their young age, primary school students may not have enough experience or

maturity to assess themselves critically (Self-assessment in 6.2.4, p.**Error! Bookmark not defined.**). Thus, guidance from teachers or in the textbook would be extremely helpful. Although peer-assessment and self-assessment were not very often used in the classroom, they are strategies that merit more use. Not only do they alleviate teachers' heavy workload, they are also independent learning habits that can benefit students for a much longer time (5.3.5).

Self-assessment, as the most popular FA strategy for the students in this study, can motivate students effectively. Self-assessment can work better if it is introduced to students when they are young, because older children become more conservative when they realise the potential social cost (Butler, 2018). To use self-assessment at an early stage in language learning can build a learner-centred learning habit which benefit students in the long term.

To improve the effectiveness of self-assessment in boosting achievement, there could be clear and specific criteria for children, which were found to be deficient in the classroom in this study and in other studies (*e.g.*: Antoniou & James, 2014). This is because self-assessment, as well as peer-assessment, develops the students' autonomy when they are aware of the assessment criteria or mark scheme (Harris, 2007). In self-assessment, the teachers' role is to provide structures of learning aims, rubrics and exemplar responses at each level, to provide time for students to internalise the success criteria, and to provide immediate feedback to support students' self-assessment (Wylie & Lyon, 2015). When a teacher guides the students to use self-assessment, it is essential to indicate which aspects they could try to assess themselves, since younger children may not understand what self-assessment is or how to do it (Black, 2015).

Self-assessment was not only a strategy to improve English learning for student, but also a strategy to improve English teaching for teachers. Teachers in China seem to have limited knowledge in self-assessment. In this study, teachers either focused on the ready-made self-assessment task in the textbook, or indicated that self-assessment is dichotomous: self-affirmation or self-denial. Teachers' views towards self-assessment would affect the perceived effectiveness of new self-assessment practice (Butler & Lee, 2010). Although learner-centred teaching has been promoted in China for a long time, not all teachers have implemented it properly, especially those who did not experience learner-centred learning when they were students, and those who do not believe in students' ability to learn and so kept control of students' learning. Teachers in other research have proposed, "what the teacher should do is to guide rather than control" (Xiao & Carless, 2013). Assessing oneself may provide teachers with an opportunity to reflect their teaching, and to experience the benefit that self-assessment brought them. Accordingly, it would be more convincing when the teachers give self-assessment instructions to students.

The table below summarises the most suitable FA strategies to motivate students from different groups, according to the findings. It is noteworthy that students from the lower-achieving group can benefit the most from FA, while students from medium-achieving groups potentially benefit the least (but that cannot deter teachers from exploring use of FA with these students in more engaging ways). Boys in the lower-achieving group found it easy to lose interest in teacher-led FA, as did girls from the higher-achieving group. The

positive impacts of student-led FA seemed to accumulate during the course of learning, but not for girls with lower achievement.

Table 7.1 Summary of suitable FA strategies for different groups of students

FA strategies	HA		MA		LA	
	Boys	Girls	Boys	Girls	Boys	Girls
SLA					√/X	√/√
O-Q					√/X	√/√
C-Q		√/X				X/√
CF-F		√/X				
S-F					√/X	√/√
CR-F						√/X
P-PA	√/X	X/√			√/√	√/√
N-PA	X/√				X/√	
sa	√/√	√/√			X/√	

For example √/X: this FA strategy was useful in the pre-test but not in the post-test

To sum up, peer-assessment and self-assessment are more effective in promoting higher achievers' motivation; FA in general has the most significant impacts on lower-achieving girls' achievement, and assessment input received from teachers have better effects than input from peers or from themselves; achievement and motivation are closely related in lower-achieving girls only. It is worth noting that these findings and recommendations are suggested for use by primary school teachers and practitioners in Jiangsu Province in China, but the results of this study cannot be generalised more widely than this without caution.

7.5.2 Teacher training

Teachers' knowledge grows with both training and experience, but the latter is hard to pass on (Rea-Dickins & Gardner, 2000; Liu & Xu, 2017). Although researchers can identify what works, it is the teacher who is the agent that actually makes things work (Black, 2015). Therefore, the training for teachers is particularly important due to its generalisation. Professional FA training is inadequate due to poor understanding of FA and improper implementation both in China (Huang & Luo, 2014; Gu, 2014) and other countries (Afitska, 2014b; Lam, 2015). To change the teachers' understanding and beliefs is part of the logic model for FA of English learners (Figure 2.6), because their philosophy about teaching and learning can affect the quality of implementation of FA (Carless, 2005; Heitink, et al., 2016; Butler & Lee, 2010). Therefore, better CPD is needed. It includes not only normal lecture training, but also other creative CPD training forms, such as visiting other schools, peer support and observation, and English teachers working with colleagues across the school (Jones & Little, 2018).

Society, schools and the teachers themselves need to create a healthy environment for their assessing practices. More professional development that meets teachers' needs may be promoted (Lee & Wiliam, 2005). RT had a three-month work visit to the UK, which may have brought her many new ideas in assessment. Teachers can develop their pedagogical content knowledge via sharing views and making changes together (Harrison, 2005). Besides this, the community needs to create a supportive environment for those who want to implement FA in their classroom (Lee, Mak, & Burns, 2016). In China, preparing lessons together is a routine for teachers who teach the same subject. However, most of the time, the teachers

will only prepare what to teach, rather than plans on how to teach or how to assess. With the information in this study, teachers can add FA strategies into their lesson preparation.

The implementation process needs time, professional training and practical tools (Gordon, et al., 2014). Two of the difficulties of implementing FA in China (discussed in 2.1.9) were: a) FA as a new concept to teachers is hard to digest; b) the concept of FA is vague to teachers. The application of the aforementioned FA strategies offers not only the practical tools for the teacher, but also a successful examples of FA that are embedded in daily classroom teaching in China. The teacher would not feel they have to learn something totally new. Teachers eventually will consider FA as redistribution of their time and efforts (Chen, 2015). Due to the heavy workload of being a primary school English teacher, there is value in having a guide for "do's and don'ts" in each FA strategy which are simple to understand. To enhance the practical application value of this research, a 'toolkit' of advice for implementing FA in schools, based on the findings of this study, is provided in 7.5.1. It contains a set of helpful and handy suggestions that are pertinent, direct and easy for teachers to follow.

Character-building in Confucian heritage reflects many dimensions of FA. First of all, Mianzi may be an obstacle when some veteran teachers try to improve their teaching. They are the honoured group in schools because of their experience, which may make it more difficult for them to adjust their teaching according to the new strategies. The experience is their honour but could turn into their burden because of the fear of losing Mianzi. Secondly, seeking harmony in the interactions and trying to keep a good Guanxi with peers may cause both teachers' and students' hesitant or fake responses in using FA. For example, confirmative feedback does not comment on the answers that are worthy of praise; peer-assessment is more emotionally driven rather than on the basis of academic evidence. On the other hand, the Confucian ideals can inspire a wise use of FA too. In this study, teachers use corrective feedback as a re-casting strategy that turns an individual mistake into an interesting lesson for everyone. It saves Mianzi and keeps the good Guanxi. Therefore, it is worth noting that Confucian heritage has its influence on many dimensions of FA. It is the people who use FA can make the impact positive or negative.

The implications in this study may provide teachers with a new perspective on using FA. They will be more eager to use FA if they are aware of the positive impact of FA (Yan & Cheng, 2015). Nevertheless, this requires English teachers in China to think differently about their teaching routine, structure and style. It is not easy to make such a significant change (Black, 2015). Teachers may need to give up some control in the classroom, which is a tough task for new teachers (Black, 2003; Lee & Wiliam, 2005; Harris, 2007). More support is needed to help teachers to adapt FA effectively, especially as Carless (2012) suggested (see Figure 2.5) at the personal level and environment level, from teacher to education reform. Teachers' professional development in innovative assessment, including FA, need to take place not only as top-down delivery comprising courses, workshops and one-off trainings organised by schools, but needs to include active learning and exploration by teachers (Liu & Xu, 2017).

Lastly, it is not only the in-service teachers in primary schools, but also the pre-service teachers in teacher education universities who are learning how to teach, who need the

training on the theoretical knowledge and the use of FA (Rea-Dickins & Gardner, 2000; Jones, 2014). Although it was reported that teachers have already incorporated the philosophy of assessment advocated by the NEC into their ideas, they had not yet applied it in practice due to the rigid education management system (Brown & Gao, 2015). Teachers form their assessment habits according to their earlier experiences being assessed as learners and implementing assessment as teachers (Liu & Xu, 2017). If FA can be well applied in the courses in teacher training institutes, the pre-service teachers would naturally apply FA in their own teaching practice. Thereafter, more teachers with this assessment philosophy would join the team of English teaching and the rigid ideas would be more likely to change. Unfortunately, such training about assessment is currently inadequate according to my personal experience as a student teacher in China.

7.6 Where to improve

The study's limitations in terms of methodology have been discussed in 3.8 and steps to eliminate or reduce those limitations are suggested. There is no such thing as perfection, including in research, but there is room for improvement in future work, in terms of the research design and the data gathering methods, and in terms of what studies can bring for the future of FA in China. These ideas will be presented in the below.

7.6.1 Limitations

Length of fieldwork period

The outcomes of the investigation of FA strategies in this study (see Figure 4.11) are consistent with Wylie and Lyon's findings (2015) that sharing learning aims, questioning and feedback were more often used than peer-assessment and self-assessment. Meanwhile, there was greater use of FA strategies in the post-test than in the pre-test. This increase over the period of the study was not as significant as in Wylie and Lyon's work. This might due to the comparative length of fieldwork: Wylie and Lyon conducted their project over two years, while this study was only for ten weeks. In such a short-term study, it was difficult to evaluate sustainable achievement. There is some promise of more significant impact if the current project could run for longer. The original plan was to conduct the fieldwork for three months, however, due to the unforeseen delays in the ethics application process, the main study of the fieldwork was forced to shrink to ten weeks.

Even though this time period was short, the procedure implemented was complete. From the successful ethics application to the well-organised field trip, from the pilot study to the main study, and from the pre-test to the post-test, every essential step in the research played its role in the rigour of the study. Although the findings need replication and extension with a much larger sample and a longer time in the field, this study emphasises the possibility and effectiveness of applying FA in English classes in China.

Size of participant sample

In this study, the sample size was small. There were twelve student interviewees, two teacher interviewees, and 196 students in the current study. Only one public school was involved. The diversity of the participants and the characteristics of the setting determine the extent to which the study's findings can be generalised. Although the sample itself had degrees of representativeness, such as different achievement levels, evenly distributed gender proportions, and different FA environments, the sample ideally needs to be extended to a larger context. There are many other types of schools in China, ranging from private schools where the English language is used to teach different subjects, and public schools in rural areas where one teacher has to teach all the subjects. The English textbooks are very different in different provinces in China. The current study was conducted in Jiangsu Province where students use the version of Jiangsu Education. Students in other provinces may use versions of People's Education, Zhejiang Education, Shanghai Education, *etc*. Therefore, the findings in this study will be most appropriate for Chinese students who are in public schools in Jiangsu Province. It is a province with a population of 79,500,000 (in 2013), and 5,282,100 primary school students (the most recent data was in 2003) (Jiangsu, 2005), so there are clearly many students and teachers who can benefit from this study.

Depth of implementation

It is worth studying each individual strategy more intensively in future work because there are many details that have emerged which merit investigation. For example, with questioning: wait time after questions is an important topic that was not within the scope of the research objectives of this study. After questions are posed, learners need time to think before giving responses (Black, 2003). Also, students are encouraged to raise questions themselves. Although this tends to slow down the pace of lessons (Harris, 2007), it is helpful to develop learner autonomy and sustainable learning. Take self-assessment as another example. The implementation of self-assessment can be differentiated more finely: as generic self-assessment and after-task self-assessment, representing semantic/ decontextualized meaning and pragmatic/contextualized meaning respectively (Butler, 2018). These different aspects of self-assessment, what students think and how they process self-assessment were not discussed in this thesis.

Rather than an exploration of different components in each FA strategy, the current study's wider lens was on the flow of FA, which is the integration of the strategies. In other words, although there is some risk of over-simplification, this study focused on the basic implementation structure for FA in primary school English classes. The deeper investigation of each strategy will be the next step in future work.

Accuracy of data

The data gathering process was not without flaws and there are improvements which could be made. For example, regarding the interviews, the length of time for interview could have been longer and the environment was not quiet all the time. Improving these factors may need more support from the school and teachers, as well as funding institutes. With the support of the teacher, students can have a special allocated time to be interviewed; with the support from the school, the interview can take place in an appropriate quiet environment. The support from teachers and schools needs not only the *Guanxi*, but also

reward incentives, such as compensation for their efforts. However, due to the researcher's limited resources, financial resources in particular, the interviews were conducted under the best circumstances they could be. Even with the constraints, the interview method was still the best way to gather this sort of data because it gave pupils a chance to speak up, rather than being observed by outsiders (in the cultural sense) who may misinterpret their behaviours. Secondly, the questionnaire relied on self-report results. The single source of their motivation makes the data authentic, but plausible. It was not only the self-reporting data from students, but also that from teachers that may influence the research findings. However, the systematic analysis of classroom observation data is time-consuming. The greater workload typically requires more researchers to be involved or a longer research time-frame to analyse the data, and these options were not viable in the current study. Lastly, the quiz was not a standardised test. Although all the questions on the quiz sheets were adapted from the teacher's reference book and were closely related to the teaching material, the difficulty level may have varied quiz to quiz. With the benefit of a longer research period, more standardised exams designed by professionals could be applied in future work to record students' achievement in an even more reliable way.

This study cannot prove that the changes in motivation and achievement are conclusively caused by FA and only FA. It is impossible to "prove with 100% certainty that a certain impact is caused by a specific intervention" in real classrooms (Andersson & Palm, 2017). Therefore, the findings pertain to the evidence of the *relationship* between FA, motivation and achievement, and how FA works from students' and teachers' viewpoints. This study provides the stepping stone toward the probability of the causal relationship that needs to be further evidenced in the future.

Even with the limitations of the present study, the research design demonstrates a thoughtful and well-planned research procedure; the data collection accords with the scientific research design; the data contributes persuasive evidence of FA's impact on students' motivation and achievement in order to provide practical suggestions to English teachers in China. Although the study was not perfect, its methods and analyses were reasonably well-designed and executed.

7.6.2 Reflection

FA in China, especially in Chinese primary schools, has a long way to go. Because of the long history of competitive examinations, the large population which may reinforce competition as a guiding principle, as well as the societal emphases on academic achievement as a measure of an individual's worth (Xiao & Carless, 2013; Carless & Lam, 2014), SA still has the more powerful role in daily teaching in China. This may be one of the reasons that implementation of FA is not yet sufficient and effective in China (Peng, 2017). Such problems are faced not only by teachers in China, but teachers in other Confucianism cultures too (Pham & Renshaw, 2015). Therefore, what can this study bring to English teachers in China?

First of all, this study does not wish to overturn the use of SA such as National College Entrance Examination (NCEE) which includes English as a compulsory exam subject. Although the NCEE is criticised by teachers and students, scholars and researchers, it is still a

relatively fair opportunity for everyone. A good example is Jiang He, who gave a speech at Harvard University's graduation ceremony in 2016 (Harvard, 2016). He came from quite a small, poor village in China. Without China's public education and NCEE, it was unlikely he would have made such a great achievement, not to mention be nominated in the '*30 under 30*' in Healthcare by Forbes in 2017. From the government's perspective, during the 40 years of the NCEE's resumption (1977-2017), many attempts were made to establish a modern examination system composed of standardised tests, comprehensive evaluation and assorted admissions criteria (OECD, 2016). NCEE is not the same NCEE it used to be. It provides more opportunities for everyone to live a better life, especially students who come from disadvantaged families. Therefore, it is not viable or desirable to completely eliminate the NCEE. It was claimed that FA implementation in China would not be fruitful unless the emphasis of high-stakes examinations could be changed (Liu & Xu, 2017). Nevertheless, this study sought to ascertain whether students' language learning can be improved via formative assessment in such an overarching SA-focused environment. The evidence for academic improvement in this study was patchy, therefore we cannot be blindly optimistic. Inferences can only be made that FA is a 'seed' with potential to improve students' achievement, but FA implementation needs further development by teachers, schools, even the education system in China (Liu & Xu, 2017) before it can realise this potential.

Secondly, changes must be made, not only in the educational reform, but also in the teachers' awareness of using FA. If teachers, as the key agents of teaching and assessing, understand the changes that have happened in the educational reform, they will believe that the purpose of education is not just success in the NCEE. If they use FA mechanically without a deep understanding, it could have a negative impact (Harlen, 2007). As Carless's framework (Figure 2.5) describes, innovation in assessment is in the personal domain, which includes teachers' understandings and beliefs. To release them from the traditional assessing habits is actually the priority. Improvement in teachers' understanding, knowledge and skills are the desired changes during FA implementation (Alvarez, Ananda, Walqui, Sato, & Rabinowitz, 2014). Of course, beliefs are not always easy to change. Substantial support would likely be essential for teachers (Andersson & Palm, 2017). Although some FA policies have been published and promoted in China, the implementation was not ideal (2.1.9). It needs a great deal of trials, successful examples, and empirical evidence. Although some teachers claimed that "we cannot change much unless the exams change" (Yan, 2015), the study suggested that FA was powerful, particularly for girls with lower achievement, in an authentic setting in an exam-oriented environment. Furthermore, different FA strategies had varied impacts on students, but in general there was benefit. Therefore, this study achieved one of its intentions, to show that assessment *'for'* learning and *'of'* learning can co-exist in harmony. This helps teachers to build confidence in FA, even in an exam-oriented environment.

Thirdly, this FA study is not a jigsaw of five separate strategies, but a synthesis. In the classroom, all FA strategies contributed to the impact of FA on students' motivation and to a lesser extent achievement. Sharing learning aims is embedded in each strategy, and questioning and feedback potentially train leaners to self-assess and peer-assess (Afitska, 2014b). My ultimate vision for English classes in China is one where students are helped to become independent, responsible and effective learners via formative assessment. Although each strategy was analysed and discussed individually, it was not my intention to

divide them, but to break down the FA into details. It does not mean that teachers should uniformly use each strategy with every student. On the contrary, this was one of the differences between the current study and other research. Each FA strategy was discussed in detail for specific target groups, rather than combining strategies together in a vague way.

However, it was observed that teachers might not use FA due to their traditional assessment habits, the perceived extra workload of FA, and the lack of empirical evidence of the positive impact of FA. All these factors led to less belief in FA and less implementation. In China, the importance of the NCEE suggests that the long-standing examination-oriented assessment culture will continue to prevail. Hence, the best and most feasible change that I can make, as a researcher, is to propose recommendations to help improve the learner experience during the learning process, by highlighting best FA practices emerging from the empirical evidence of this study. Therefore, the following section sets out the implications for each FA strategy along with suggestions for training for teachers.

7.7 Where to next?

The current project has contributed to the practical application of FA in the context of the English language learning classroom in China. According to the fieldwork, findings and discussions of this study, there is further innovation in FA which could be developed and researched.

7.7.1 Teachers' and students' personalities

In the current study, many students and both teachers reflected that the teacher's personality and teaching style would influence the FA implementation. According to LT, she had a so-called 'aloof' style, whereas RT was more demonstrably affectionate with her students. Teachers with different personalities and styles will create slightly different classroom environments. Social norms in class and subject-specific norms will be different, meanwhile, they are highlighted in effective classroom environments (Boaler & Humphreys, 2005; McClain & Cobb, 2001). FA strategies can be seen, in a way, as a part of the norms in different classrooms. Therefore, the effectiveness of FA in a class may affected by teachers' personalities and approaches.

Students' personalities are closely related to their academic motivation and achievement (Komarraju, Karau, & Schmeck., 2009; Clark & Schroth, 2010). Their learning styles also can affect achievement (Komarraju, Karau, Schmeck, & Avdic, 2011) and this may especially be reflected in peer-assessment and self-assessment. Teachers usually know their students well so it may be possible for teachers to make FA more effective by tailoring it to students' personalities and learning styles.

The personal relationship between teacher and students, or among students, is one of the key factors influencing the effectiveness of FA (Crooks, 2001). For example, students should be encouraged to express their understanding and opinions at any time during the lesson (Black & Wiliam, 1998b). However, this suggestion conflicts with Chinese classroom traditions, where students listen to the teachers in the class and ask question afterwards. If the relationship between teacher and students is less strict and hierarchical, students may

feel free to ask questions in the class, and accordingly engage in the teaching and learning more.

In sum, the personalities, teaching and learning styles of teachers and students may affect the relationship between individuals and influence the use and effectiveness of FA. Therefore, future study might explore these aspects of FA dynamics. Case studies in specific contexts may be the most appropriate approach, considering the heavily socially determined nature of the relationships.

7.7.2 Multiple assessment

FA is not the only assessment strategy in the class nor necessarily the most appropriate (Black & Wiliam, 2009). The guidelines in the latest educational reforms in China stress the 'Core Competencies', which aim at achieving the goal of 'all-round development of human beings' (The State Council Information Office of the People's Repubic of China, 2016). The core competencies of our students mainly related to three categories and six aspects, namely: the cultural foundation (Humanities Spirit, Science Spirit); independent development (learn how to learn, healthy living) and social participation (responsibility, practical innovation). The new guidelines seek to promote a better environment where FA or other forms of assessment can be applied. As mentioned in 2.1.3, schools in China will follow the guidance closely, due to the top-down nature and weight of authority in policy-making in this context. It is not only the policy being handed down, but all the studies and researchers' efforts underpinning the policy. Since the latest policy asserts the importance of core competencies, there will be a strong need for multiple assessments in primary education. It is no longer just about students' academic achievement, but the ability in students to solve problems, particularly problems in reality or simulations of reality. New assessment styles, such as performance assessment, is focused on the assessment of students' abilities. They can be combined with formative assessment and summative assessment to make a difference in the changing trends for assessment in China. Although it seems utilitarian, it is how assessment in schools may develop in the future in China.

7.7.3 ICT-based assessment

In traditional classes, teachers take most of the responsibility for the assessment, which increases their workload hugely (Jonsson, Lundahl, & Holmgren, 2015). This may hinder teachers from using certain effective means of assessment. In addition to that, the large number of children in a typical class in China (2.1.9) makes it even more difficult to implement formative assessment. Some interviewees also mentioned that the influence that technology may bring to their learning (6.5.2). Therefore, ICT-based assessment offers a convenient alternative method for teachers to reduce their workload, and for students to enrich their learning environment.

New technology is developing everywhere in education. There are some successful examples of using FA with technology. Similar to Gordon and his colleagues' successful computer-assisted FA (2014), many other researchers also showed that compared to assessment by teachers only, technology-based tools provide more dynamic feedback to young children which can improve their achievement and motivation (Tzuriel & Shamir,

2002; Faber, Luyten, & Visscher, 2017; Ponce, Mayer, Figueroa, & López, 2018). In the current study, the teacher did use computers in teaching, but only for the purposes of presenting information to be learned. It is worth encouraging the teachers to try something new but effective for assessment purposes.

The participating schools in this study are able to use existing facilities to provide ICT-based assessment. Most of the schools in cities in Jiangsu Province can afford this in terms of technology and cost. Taking another city in Jiangsu as an example, in Suzhou, 220 of the city's public schools have computers in each classroom, with 100Mbps broadband since 2013 (Jiangsu, 2016). Therefore, ICT-based assessment is a feasible method to promote FA in school. The main drawback is that schools in the countryside are not currently equipped to benefit from ICT-based assessment. For those schools which cannot afford essential equipment and services (and these are not a small minority), it is even more important to use traditional 'low-tech' FA strategies discussed in this study.

7.7.4 Beyond classroom assessment

In the current study, all the assessment strategies took place in the classroom. However, students mentioned many other factors beyond the classroom that can motivate their learning and improve their achievement, such as the input of parents (mentioned many times by interviewees) or the impact of family status. In China, parents from higher socio-economic backgrounds have shown more support for children's learning autonomy and less control over their study (Bulter, 2014). The nurturing style is related to children's personality, even their learning style. It would be useful if the relationships between these factors are investigated. Interviewees also mentioned the games in English language learning. Most children in China are equipped with a mobile phone. The opportunities for exposure to mobile games are increasing rapidly. Using games via mobile phones to assess students' English learning has potential to motivate students to the greatest extent. Two of the advantages are that children may perceive 'assessment' no longer as just a serious thing in the classroom, and the data can be easily collected electronically for analysis in order to adjust of teaching and learning accordingly.

In contribution to the debate about FA's influence on achievement, this study has started to define an age-appropriate and creditable assessment approach for research, and contributed to the literature around FA's positive impact on young students' achievement and their motivation in language learning.

"Every ending is a new beginning". Obviously, this is not the end, and I am looking forward to continuing my study on formative assessment.

Bibliography

Afitska, O. (2014a). Formative Use of Language Assessment: Supporting Learning Through the Medium of a Second Language. *US-China Education Review A, 4*(8), 517-535.

Afitska, O. (2014b). 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), 29-39.

Afitska, O. (2015). Role of focus-on-form instruction, corrective feedback and uptake in second language classrooms: some insights from recent second language acquisition research. *The Language Learning Journal, 43*(1), 57-73.

Airasian, P. (2001). *Classroom Assessment: Concepts and Applications* (4 ed.). Boston: McGraw-Hill.

Akomolafe, M. J., Ogunmakin, A. O., & Fasooto, G. M. (2013). The role of academic self-efficacy, academic motivation and academic self-concept in predicting secondary school students' academic performance. *Journal of Educational and Social Research, 3*(2), 335-342.

Alvarez, L., Ananda, S., Walqui, A., Sato, E., & Rabinowitz, S. (2014). *Focusing formative assessment on the needs of English language learners.* San Francisco, CA: WestEd.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of educational psychology, 84*(3), 261-271.

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.

Andrade, H. L., Wang, X., Du, Y., & Akawi, R. L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *The Journal of Educational Research, 102*(4), 287-302.

Antoniou, P., & James, M. (2014). Exploring formative assessment in primary school classrooms: Developing a framework of actions and strategies. *Educational Assessment, Evaluation and Accountability, 26*(2), 153-176.

Association, A. E. (1999). *American Educational Research Association, American Psychological Association, and National Council on Measurement in Education.* Standards for educational and psychological testing. Washington, D.C.: American Educational Research Association.

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological review, 64*(6), 359-372.

Atkinson, J. W. (1964). *An introduction to motivation.* Oxford, England: Van Nostrand.

Atkinson, J. W., Raynor, J. O., & Birch, D. (1974). *Motivation and achievement.* Washington DC: Winston.

Bachman, L., & Palmer, A. (1996). *Language Testing in Practice.* Oxford: Oxford University Press.

Bakker, J. T., Denessen, E., Bosman, A. M., Krijger, E. M., & Bouts, L. (2007). Sociometric status and self-image of children with specific and general learning disabilities in Dutch general and special education classes. *Learning Disability Quarterly, 31*(1), 47-62.

Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1988). Self-Regulation of Motivation and Action Through Goal Systems. In H. V., B. G.H., & F. N.H. (Eds.), *Cognitive Perspectives on Emotion and Motivation. NATO ASI Series (Series D: Behavioural and Social Sciences)* (Vol. 44, pp. 37-61). Dordrecht: Springer.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: W. H. Freeman.

Banks, M. (2007). *Using visual data in qualitative research.* London: SAGE Publications Ltd.

Baron-Cohen, S. (2004). *The essential difference.* UK: Penguin.

Barratt, R. (1998). Shaping middle schooling in Australia: A report of the National Middle Schooling Project. *Australian Curriculum Studies Association*.

Bazeley, P., & Jackson, K. (2013). *Qualitative data analysis with NVivo* (2 ed.). London: SAGE Publications Ltd.

Bell, B., & Cowie, B. (2002). *Formative Assessment and Science Education.* New York: Kluwer Academic Publishers.

Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5-25.

Black, P. (1995). Can teachers use assessment to improve learning? *British Journal of Curriculum & Assessment, 5*(2), 7-11.

Black, P. (1998). *Testing, friend or foe?: the theory and practice of assessment and testing.* London: Falmer Press.

Black, P. (2003). The nature and value of formative assessment for learning. *Improving schools, 6*(3), 7-22.

Black, P. (2015). Formative assessment–an optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice, 22*(1), 161-177.

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.* London: School of Education King's College.

Black, P., & Wiliam, D. (2004). The formative purpose: Assessment must first promote learning. *Yearbook of the National Society for the Study of Education, 103*(2), 20-50.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*, 5-31.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). *Working inside the black box: Assessment for learning in the classroom.* London: School of Education, King's College.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice.* Buckingham: Open University Press.

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment, 1*(2), 1-12.

Bloom, B. S. (1968). *Toward A Theory of Testing Which Includes Measurement-Evaluation-Assessment.* Center for the Study of Evaluation, Instructional Programs Occasional Report No. Nine.

Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. In R. W. Tyler (Ed.), *Education evaluation: New roles, new means (National Society for the Study of Education Yearbook)* (Vol. 68, pp. 26-50). Chicago, IL: University of Chicago Press.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning.* New York: McGraw-Hill.

Boaler, J., & Humphreys, C. (2005). *Connecting mathematical ideas: Middle school video cases to support teaching and learning.* Portsmouth: Heinemann.

Broussard, S. C., & Garrison, M. E. (2004). The relationship between classroom motivation and academic achievement in elementary-school-aged children. *Family and Consumer Sciences Research Journal, 33*(2), 106-120.

Brown, G. T., & Gao, L. (2015). Chinese teachers' conceptions of assessment for and of learning: Six competing and complementary purposes. *Cogent Education, 2*(1), 1-19.

Brown, G. T., Kennedy, K. J., Fok, P. K., Chan, J. K., & Yu, W. M. (2009). Assessment for student improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in education: principles, policy & practice, 16*(3), 347-363.

Brown, H. D. (1981). Affective factors in second language learning. In J. Alatis, H. Altman, & P. Alatis (Eds.), *The Second Language Classroom: Directions for the Eighties* (pp. 111-129). New York: Oxford University Press.

Brown, H. D. (1990). M & Ms for language classrooms? Another look at motivation. *Georgetown University round table on language and linguistics*, 383-393.

Brown, H. D. (1994). *Teaching by Principles.* Englewood Cliffs, NJ: Prentice Hall.

Brown, S., & Mcintyre, D. (1993). *Making Sense of Teaching.* Phiadelphia, PA: Open University Press.

Bryman, A. (2016). *Social Reserach Methods* (5 ed.). Oxford: Oxford University Press .

Bryman, A., & Cramer, D. (2011). *Quantitative data analysis with IBM SPSS 17, 18 & 19: A guide for social scientists.* London: Routledge.

Buchner, E. F. (1904). *The Educational Theory of Immanuel Kant.* Philadelphia: Lippincott.

Bulter, Y. G. (2014). Parental factors in children's motivation for learning English: a case in China. *Research Papers in Education, 30*(2), 164-191.

Butler, R. (1989). On the psychological meaning of information about competence: A reply to Ryan and Deci's comment on Butler (1987). *Journal of Educational Psychology, 81*(2), 269-272.

Butler, Y. G. (2015). English language education among young learners in East Asia: A review of current research (2004–2014). *Language Teaching, 48*(3), 303-342.

Butler, Y. G. (2018). The Role of Context in Young Learners' Processes for Responding to Self-Assessment Items. *The Modern Language Journal, 102*(1), 242-261.

Butler, Y. G., & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing, 27*(1), 5-31.

Butler, Y. G., & Zeng, W. (2014). Young Foreign Language Learners' Interactions During Task-Based Paired Assessments. *Language Assessment Quarterly, 11*(1), 45-75.

Cai, H., Kwan, V. S., & Sedikides, C. (2012). A sociocultural approach to narcissism: The case of modern China. *European Journal of Personality, 26*(5), 529-535.

Cameron, L. (2001). *Teaching Language to Young Learners.* Cambridge: Cambridge University Press.

Cameron, L., Erkal, N., Gangadharan, L., & Meng, X. (2013). Little Emperors: Behavioral Impacts of China's One-Child Policy. *Science, 339*(6122), 953-957.

Carless, D. (2005). Prospects for the implementation of assessment for learning. *Assessment in Education: Principles, Policy & Practice, 21*, 5-13.

Carless, D. (2007). The PEAL project: co-operation or competition in the assessment process. *Paper presented at the British Educational Research Association Annual Conference, 5*, 8.

Carless, D. (2012). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings.* London: Routledge.

Carless, D., & Lam, R. (2014). Developing Assessment for Productive Learning in Confucian-Influenced Settings: Potentials and Challenges. In C. W.-S. al. (Ed.), *Designing Assessment for Quality Learning* (pp. 167-179). Springer Netherlands.

Carreira, J. M., Ozaki, K., & Maeda, T. (2013). Motivational model of English learning among elementary school students in Japan. *System, 41*(3), 706-719.

Carrillo-de-la-Peña, M. T., Baillès, E., Caseras, X., Martínez, À., Ortet, G., & Pérez, J. (2009). Formative assessment and academic achievement in pre-graduate students of health sciences. *Advances in Health Sciences Education, 14*(1), 61-67.

Carson, J. G., & Nelson, G. L. (1996). Chinese students' perceptions of ESL peer response group interaction. *Journal of Second Language Writing, 5*(1), 1-19.

Cauley, K. M., & McMillan, J. H. (2010). Formative Assessment Techniques to Support Student Motivation and Achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 83*(1), 1-6.

Chen, C., & Stevenson, H. W. (1995). Motivation and mathematics achievement: A comparative study of Asian-American, Caucasian-American, and East Asian high school students. *Child development, 66*(4), 1215-1234.

Chen, J. (2015). Formative assessment as a vehicle for changing classroom practice in a specific cultural context. *Cultural Studies of Science Education, 10*(3), 753-762.

Chen, J., & Brown, G. T. (2018). Chinese secondary school students' conceptions of assessment and achievement emotions: endorsed purposes lead to positive and negative feelings. *Asia Pacific Journal of Education*, 1-19.

Chen, Q. (2011). Assessment policy change in relation to English language teaching and learning in China: A study of perspectives from two universities. *PhD Dissertation*. Retrieved from http://eprints.qut.edu.au/45767/

Chen, Q., Kettle, M. A., Klenowski, V., & May, L. (2013). Interpretations of formative assessment in the teaching of English at two Chinese universities: A sociocultural perspective. *Assessment & Evaluation in Higher Education, 38*(7), 831-846.

Chen, Q., May, L., Klenowski, V., & Kettle, M. A. (2014). The enactment of formative assessment in English language classroom in two Chinese universities: teacher and student responses. *Assessment in Education: Principles, Policy & Practice, 21*(3), 271-285.

Chen, X., & Chen, C. C. (2004). On the Intricacies of the Chinese Guanxi: A Process Model of Guanxi Development. *Asia Pacific Journal of Management, 21*(3), 305–324.

Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language Testing, 22*(1), 93-121.

Cherng, H.-Y. S., Hannum, E. C., & Lu, C. (2014). China. In P. A. Stevens, & A. G. Dworkin (Eds.), *The Palgrave Handbook of Race and Ethnic Inequalities in Education* (pp. 205–37). London: Palgrave Macmillan UK.

China Education and Reserach Network. (2005, 10 18). Retrieved from Suzhi Jiaoyu Chronicle of Events, Suzhi jiaoyu dashiji: http://www.edu.cn/20051018/3156152.shtml

Chow, S. J., & Yong, B. C. (2013). Secondary School Students' Motivation and Achievement in Combined Science. *US-China Education Review B, 3*(4), 213-228.

Clapham, C. (2000). Assessment and Testing. *Annual Review of Applied Linguistics*, 147–161.

Clark, M. H., & Schroth, C. A. (2010). Examining relationships between academic motivation and personality among college students. *Learning and individual differences, 20*(1), 19-24.

Clough, P., & Nutbrown, C. (2007). *a student's guide to METHODOLOGY Justifying Enquiry* (2nd ed.). London: SAGE Publications Ltd.

Cohen, J. W. (1988). *Statistical power analysis for the behavioural science* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). Oxon: Routledge.

Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual review of psychology, 51*(1), 171-200.

Crooks, T. (2001). The validity of formative assessments. *British Educational Research Association Annual Conference* (pp. 13-15). Leeds: University of Leeds.

Crotty, M. (1998). *The Foundation of Social Research: Meaning and perspectives in the Research Process.* London: SAGE.

Cui, K. (2015). The insider–outsider role of a Chinese researcher doing fieldwork in China: The implications of cultural context. *Qualitative Social Work, 14*(3), 356-369.

Curry, K. A., Mwavita, M., Holter, A., & Harris, E. (2016). Getting assessment right at the classroom level: using formative assessment for decision making. *Educ Asse Eval Acc, 28*, 89-104.

Côté, S., Tremblay, R. E., Nagin, D., Zoccolillo, M., & Vitaro, F. (2002). The development of impulsivity, fearfulness, and helpfulness during childhood: Patterns of consistency and change in the trajectories of boys and girls. *Journal of Child Psychology and Psychiatry, 43*(5), 609-618.

Dann, R. (2002). *romoting assessment as learning: Improving the learning process.* London: Routledge.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behaviour.* New York: Plenum.

Deci, E. L., & Ryan, R. M. (1991). A motivational approach to self: Integration in personality. In *Perspectives on motivation: Nebraska symposium on motivation* (pp. 237-288). the University of Nebraska Press.

Deci, E. L., & Ryan, R. M. (2002). Overview of self-determination theory: An organismic dialectical perspective. *Handbook of self-determination research*, 3-33.

Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *Educational psychologist, 26*(3-4), 325-346.

Dello-Iacovo, B. (2009). Curriculum reform and 'Quality Education' in China: An overview. *International Journal of Educational Development*, 241–249.

Denzin, N. K. (2006). Strategies of Multiple Triangulation. In A. Bryman, *Mixed Methods* (Vol. 2, pp. 195-216). London: SAGE Publications.

Dey, I. (2005). *Qualitative data analysis: A user friendly guide for social scientists.* London: Routledge.

Dörnyei, Z. (1994). Motivation and motivating in the foreign language classroom. *The modern language journal, 78*(3), 273-284.

Dörnyei, Z. (1998a). Motivation in second and foreign language learning. *Language teaching, 31*(3), 117-135.

Dörnyei, Z. (1998b, March). What is motivation? *AAAL 1998 Conference*.

Dörnyei, Z. (1999). Motivation. In S. B (Ed.), *Concise encyclopedia of educational linguistics* (pp. 525-532). Oxford, UK: Elsevier.

Dörnyei, Z. (2001). *Motivational strategies in the language classroom.* Cambridge, England: Cambridge University Press.

Dörnyei, Z. (2003). Attitudes, orientations, and motivations in language learning: Advances in theory, research, and applications. *Language learning, 53*(S1), 3-32.

Dörnyei, Z., & Ottó, I. (1998). Motivation in action: A process model of L2 motivation. In *Working Papers in Applied Linguistics* (pp. 43-69). Thames Valley University.

Dörnyei, Z., & Ushioda, E. (2013). *Teaching and researching: Motivation.* Routledge.

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

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*(10), 1040-1048.

Dweck, C. S. (1991). Self-theories and goals: Their role in motivation, personality, and development. In *Perspective on Motivation: Nebraska symposium on motivation* (pp. 199-235). the University of Nebraska Press.

Easterby-Smith, M., Thorpe, R., & Lowe, A. (2002). *Management Research: An Introduction* (2nd ed.). London: SAGE.

Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of women quarterly, 11*(2), 135-172.

Eccles, J. S. (2005). Subjective task value and the Eccles et al. model of achievement-related choices. In A. J. Elliot, & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 105-121). New York: The Guilford Press.

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and social psychology bulletin, 21*(3), 215-225.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual review of psychology, 53*(1), 109-132.

Eccles-Parsons, J. E., & Goff, S. B. (1978). Achievement motivation: A dual modality. *Educational Psychologist, 13*(1), 93-96.

Eccles-Parsons, J. E., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviours. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75-146). San Francisco, CA: Freeman.

Elig, T. W., & Frieze, I. H. (1980). Measuring causal attributions for success and failure. In F. L.J. (Ed.), *Achievement Motivation* (pp. 289-311). Boston, MA: Springer.

Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational psychologist, 34*(3), 169-189.

Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of personality and social psychology, 72*(1), 218-232.

Elliot, A. J., & Thrash, T. M. (2001). Achievement goals and the hierarchical model of achievement motivation. *Educational Psychology Review , 13*(2), 139-156.

Ellis, L. (2008). *Sex differences: Summarizing more than a century of scientific research.* Taylor & Francis.

Faber, J. M., Luyten, H., & Visscher, A. J. (2017). The effects of a digital formative assessment tool on mathematics achievement and student motivation: Results of a randomized experiment. *Computers & education, 106*, 83-96.

Fan, X., & Chen, M. (2001). Parental Involvement and Students' Academic Achievement: A Meta-Analysis. *Educational Psychology Review, 13*(1), 1-22.

Flick, U. (2014). *An introduction to qualitative research* (5th ed.). London: SAGE.

Fortier, M. S., Vallerand, R. J., & Guay, F. (1995). Academic motivation and school performance: Toward a structural model. *Contemporary educational psychology, 20*(3), 257-274.

Foster, V. (2000). Gender, schooling achievement and post-school pathways: beyond statistics and populist discourse. In T. Maxwell (Ed.), *Teaching in Context.* Canberra.

Fröjd, S. A., Nissinen, E. S., Pelkonen, M. U., Marttunen, M. J., Koivisto, A. M., & Kaltiala-Heino, R. (2008). Depression and school performance in middle adolescent boys and girls. *Journal of Adolescence, 31*, 485-498.

Frieze, I. H. (1976). Causal attributions and information seeking to explain success and failure. *Journal of Research in Personality, 10*(3), 293-305.

Frieze, I. H., & Snyder, H. N. (1980). Children's beliefs about the causes of success and failure in school settings. *Journal of Educational Psychology, 72*(2), 186-196.

Fu, R., Chen, X., Wang, L., & Yang, F. (2016). Development Trajectories of Academic Achievement in Chinese Children: Contributions of Early Scocial Behavioural Functioning. *Journal of Educational Psychology, 108*(7), 1001-1012.

Fuller, F., & Bown, O. (1975). Becoming a Teacher. In K. Ryan (Ed.), *Teacher Education: Seventy-Fourth Yearbook of the National Society for the Study of Education.* Chicago: University of Chicago Press.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation.* London: Edward Arnold.

Gardner, R. C. (2001). Integrative motivation and second language acquisition. In Z. Dörnyei, & R. Schmidt (Eds.), *Motivation and second language acquisition* (pp. 1-20). Honolulu, HI: University of Hawai'i Press.

Gardner, R. C., Tremblay, P. F., & Masgoret, A.-M. (1997). Towards a full model of second language learning: An empirical investigation. *The Modern Language Journal*, 344-362.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston, MA: Allyn & Bacon.

Goodenow, C. (1993). Classroom belonging among early adolescent students: Relationships to motivation and achievement. *The Journal of Early Adolescence, 13*(1), 21-43.

Gordon, E. W., McGill, M. V., Sands, D. I., Kalinich, K. M., Pellegrino, J. W., & Chatterji, M. (2014). Bringing formative classroom assessment to schools and making it count. *Quality Assurance in Education, 22*(4), 339-352.

Graham, S. (1991). A review of attribution theory in achievement contexts. *Educational Psychology Review, 3*(1), 5-39.

Graham, S., & Weiner, B. (1996). Theories and principles of motivation. *Handbook of educational psychology, 4*, 63-84.

Graham, S., & Williams, C. (2009). An attributional approach to motivation in school. In *Handbook of motivation at school* (pp. 11-34). New York: Roultedge.

Gray, D. E. (2014). *Doing Research in the Real World* (3rd ed.). London: SAGE.

Greig, A., Taylor, J., & MacKay, T. (2007). Introduction to research and children: a special relationship. In *Doing research with children* (pp. 2-15). London: SAGE Publications Ltd.

Greig, A., Taylor, J., & MacKay, T. (2012). *Doing research with children A practical guide* (3 ed.). SAGE.

Gu, P. Y. (2014). The unbearable lightness of the curriculum: what drives the assessment practices of a teacher of English as a Foreign Language in a Chinese secondary school? *Assessment in Education: Principles, Policy & Practice, 21*(3), 286-305.

Guan, S., & Ploner, J. (2018). The influence of cultural capital and mianzi (face) on mature students' orientation towards higher education in China. *Compare: A journal of Comparative and International Education*, 1-17.

Guoxing, Y., & Yan, J. (2014). English language assessment in China: policies, practices and impacts. *Assessment in Education: Principles, Policy & Practice* , 245-250.

Habibis, D. (2010). Ethics and Social Research . In M. Walter (Ed.), *Social Research Methods* (2nd Edition ed., pp. 89-122). Oxford: Oxford University Press.

Hainey, T., Connolly, T. M., Boyle, E. A., Wilson, A., & Razak, A. (2016). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education, 102*, 202-223.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological science in the public interest, 8*(1), 1-51.

Hammond, M., & Wellington, J. (2013). *Research Methods The Key Concepts.* London: Routledge.

Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research, 52*(1), 123-142.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (1998). Does special education raise academic achievement for students with disabilities?

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of applied econometrics, 18*(5), 527-544.

Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The Market for Teacher Quality.* Washington. DC: National Bureau of Economic Research.

Harlen, W. (2007). Criteria for evaluating systems for student assessment. *Studies in Educational Evaluation, 33*(1), 15-28.

Harris, L. (2007). Employing formative assessment in the classroom. *Improving Schools*, 249-260.

Harris, L. R., & Brown, G. T. (2013). 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.

Harris, L. R., Brown, G. T., & Harnett, J. A. (2014). Understanding classroom feedback practices: A study of New Zealand student experiences, perceptions, and emotional responses. *Educational Assessment, Evaluation and Accountability, 26*(2), 107-133.

Harris, L. R., Brown, G. T., & Harnett, J. A. (2015). Analysis of New Zealand primary and secondary student peer-and self-assessment comments: Applying Hattie and Timperley's feedback model. *Assessment in Education: Principles, Policy & Practice, 22*(2), 265-281.

Harrison, C. (2005). Teachers developing assessment for learning: Mapping teacher change. *Teacher Development, 9*(2), 255-263.

Harvard, U. (2016, May 26). *Graduate Speaker Jiang He | Harvard Commencement 2016.* Retrieved from Youtube: https://www.youtube.com/watch?v=BM2HMoK5aIU

Hattie, J. (1999). *Influences on student learning.* New Zealand: University of Auckland, Inaugural professorial lecture.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research, 77*(1), 81-112.

Heider, F. (1958). *The psychology of interpersonal relations.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Heitink, M. C., Van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational research review, 17*, 50-62.

Henly, D. C. (2003). Use of Web-based formative assessment to support student learning in a metabolism/nutrition unit. *European Journal of Dental Education, 7*(3), 116-122.

Heritage, M. (2009a). The Process of formative assessment. *Iowa Assessment for Learning Institute.* IA: Des Monies.

Heritage, M. (2009b). Understanding formative assessment and utilizing it to improve Classroom instruction. *REL Midwest at Learning Point Associates' Lessons Learned about Formative Assessment Use.* Chicago, IL.

Heritage, M. (2010). *Formative assessment: Making it happen in the classroom.* Thousand Oaks, CA: Corwin Press.

Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practice, 28*(3), 24-31.

Ho, D. Y. (1986). Chinese pattern of socialization: A critical review. In M. H. Bond (Ed.), *The psychology of the Chinese people* (pp. 1-37). New York, NY: Oxford University Press.

Hsieh, P., Sullivan, J. R., & Guerra, N. S. (2007). A closer look at college students: Self-efficacy and goal orientation. *Journal of Advanced Academics, 18*(3), 454-476.

Hu, G. (2002). Potential cultural resistance to pedagogical imports: The case of communicative language teaching in China. *Language Culture and Curriculum, 15*(2), 3-105.

Huang, J., & Luo, S. (2014). Formative Assessment in L2 Classroom in China: the Current Situation, Predicament and Future. *Indonesian Journal of Applied Linguistics, 3*(2), 18-34.

Huinker, D., & Freckmann, J. (2009). Linking principles of formative assessment to classroom practice. *Wisconsin Teacher of Mathematics, 60*(2), 6-11.

Huitt, B., Hummel, J., & Kaeck, D. (2001). Assessment, measurement, evaluation, and research. *Educational Psychology Interactive*. Retrieved from http://www.adprima.com/measurement.htm

Hvistendahl, M. (2013). Making a selfish generation by fiat. *Science, 339*(6116), 131-131.

Institute of Education Sciences, U. D. (2010). *Educational Testing Service*. Retrieved from English Learner Formative Assessment (ELFA) Project: https://www.ets.org/research/topics/ella/elfa

Jacobs, L., Gao, G., & Herbig, P. (1995). Confucian roots in China: a force for today' s business. *Management Decision, 3*(10), 29-34.

Jiang, Y. (2014). Exploring teacher questioning as a formative assessment strategy. *RELC Journal, 45*(3), 287-304.

Jiangsu, B. o. (2005, Feb 08). *Statistical bulletin of the national economic and social development of jiangsu province*. Retrieved from Bureau of Statistics of Jiangsu: http://www.jssb.gov.cn/tjxxgk/xwyfb/tjgbfb/sjgb/200502/t20050208_6130.html

Jiangsu, E. D. (2016, 12 20). *Jiangsu education brief report.* Retrieved from Education Department of Jiangsu Province: http://www.ec.js.edu.cn/art/2016/12/30/art_4381_204071.html

Jick, T. D. (2006). Mixing Qualitative and Quantitative Methods: Triangulation in Action. In A. Bryman, *Mixed Methods* (Vol. 2, pp. 217-228). London: SAGE Publications.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research , 1*(2), 112-133.

Jones, J. (2010). The role of assessment for learning in the management of primary to secondary transition: Implications for language teachers. *The Language Learning Journal, 38*(2), 175-191.

Jones, J. (2014). Student teachers developing a critical understanding of formative assessment in the modern foreign languages classroom on an initial teacher education course. *The Language Learning Journal, 42*(3), 275-288.

Jones, J. (2015). Student Teachers' Understanding of Achievement: An Exploration. *Education Today, 65*(2), 18-25.

Jones, J., & Coffey, S. (2017). *Modern Foreign Languages 5-11: a guide for teachers.* London: Routledge.

Jones, J., & Little, S. (2018). *Report of Examiners for the Degree of Doctor of Philosophy (PhD).* Sheffield : The University of Sheffield.

Jonsson, A., Lundahl, C., & Holmgren, A. (2015). Evaluating a large-scale implementation of Assessment for Learning in Sweden. *Assessment in Education: Principles, Policy & Practice, 22*(1), 104-121.

Kaplan, A., & Maehr, M. L. (2000). The Contributions and Prospects of Goal Orientation Theory. *Educational Psychology Review, 19*, 141-184.

Keblawi, F. (2006). Orientations, attitudes and de-motivation: A case study of Muslim Arab learners of English in public schools in Northern Israel. *Unpublished PhD study*.

Kelley, H. H., & Michela, J. L. (1980). Attribution theory and research. *Annual review of psychology, 30*(1), 457-501.

King, R. B. (2016). Gender differences in motivation, engagement and achievement are related to students' perceptions of peer—but not of parent or teacher—attitudes toward school. *Learning and Individual Differences, 52*, 60–71.

Kitsantas, A., Steen, S., & Huie, F. (2017). The role of self-regulated strategies and goal orientation in predicting achievement of elementary school children. *International Electronic Journal of Elementary Education, 2*(1), 65-81.

Kizlik, B. (2017, March 11). *Measurement, Assessment, and Evaluation in Education*. Retrieved from http://www.adprima.com/measurement.htm

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254-284.

Komarraju, M., Karau, S. J., & Schmeck., R. R. (2009). Role of the Big Five personality traits in predicting college students' academic motivation and achievement. *Learning and individual differences, 19*, 47-52.

Komarraju, M., Karau, S. J., Schmeck, R. R., & Avdic, A. (2011). The Big Five personality traits, learning styles, and academic achievement. *Personality and individual differences, 51*(4), 472-477.

Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing, 32*(2), 169-197.

Lam, R. (2016). Assessment as learning: examining a cycle of teaching, learning, and assessment of writing in the portfolio-based classroom. *Studies in higher education, 41*(11), 1900-1917.

Lee, C., & Wiliam, D. (2005). Studying Changes in the Practice of Two Teachers Developing Assessment for Learning. *Teacher Development, 9*(2), 265-283.

Lee, I., Mak, P., & Burns, A. (2016). EFL teachers' attempts at feedback innovation in the writing classroom. *Language teaching research, 20*(2), 248-269.

Lewin, C. (2005). Elementary quantitative methods. *Research methods in the social sciences*, 215-225.

Lewin, K., & Wang, L. (2012). University entrance examinations in China: a quiet revolution. In P. Broadfoot, R. Murphy, & H. Torrance (Eds.), *Changing Educational Assessment: International Perspectives and Trends* (pp. 153-176). New York: Routledge.

Liu, F. (2006). Boys as only-children and girls as only-children—parental gendered expectations of the only-child in the nuclear Chinese family in present-day China. *Gender and Education, 18*(5), 491-505.

Liu, J., & Xu, Y. (2017). Assessment for Learning in English Language Classrooms in China: Contexts, Problems, and Solutions. In *Innovation in Language Learning and Teaching* (pp. 17-37). London: Palgrave Macmillan.

Locke, E. A. (1996). Motivation through conscious goal setting. *Applied and preventive psychology, 5*(2), 117-124.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance.* Englewood Cliffs, NJ: Prentice Hall.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist,, 57*(9), 705–717.

Locke, E. A., & Latham, G. P. (2006). New directions in goal-setting theory. *Current directions in psychological science, 15*(5), 265-268.

Luckin, R., Clark, W., Avramides, K., Hunter, J., & Oliver, M. (2016). Using teacher inquiry to support technology-enhanced formative assessment: a review of the literature to inform a new method. *Interactive learning environments, 25*(1), 85-97.

Luo, S., Huang, J., & Xiaolei, M. (2014). *Enhancing Learning through Formative Assessment in L2 Classrooms.* Beijing: Foreign Language Teaching and Research Press.

Lysaght, Z., & O'Leary, M. (2013). An instrument to audit teachers' use of assessment for learning. *Irish Educational Studies, 32*(2), 217-232.

Machell, K. A., Blalock, D. V., Kashdan, T. B., & Yuen, M. (2016). Academic achievement at the cost of ambition: The mixed results of a supportive, interactive environment on socially anxious teenagers. *Personality and Individual Differences, 86*, 166–171.

Matthews, B., & Ross, L. (2010). *Research methods.* Essex: Pearson Education Limited.

McAlpine, M. (2002). *Principles of assessment.* Glasgow: University of Glasgow, Robert Clark Centre for Technological Education.

McClain, K., & Cobb, P. (2001). An analysis of development of sociomathematical norms in one first-grade classroom. *Journal for Research in Mathematics Education*, 236-266.

McClelland, D. C. (1967). *Achieving society.* Simon and Schuster.

McManus, S. (2008). *Attributes of effective formative assessment*. Retrieved from Council of Chief State School Officers: http://www.ccsso.org/Documents/2008/Attributes_of_Effective_2008.pdf

Ministry of Education. (2011). Retrieved from New English Curriculum for Chinese Primary Schools and Junior/Senior Middle Schoo ls: https://wenku.baidu.com/view/4426e428cfc789eb172dc82c.html

Ministry of Education, N. Z. (1994). *Assessment: Policy to Practice.* Wellington: Learning Media.

Moore, N. (2006). *How to do research A practical guide to designing and managing resarch projects* (3rd, revised ed.). London: Facet Publishing.

Morris, F. A., & Tarone, E. E. (2003). Retracted: Impact of Classroom Dynamics on the Effectiveness of Recasts in Second Language Acquisition. *Language Learning, 53*(2), 325-368.

Muola, J. M. (2010). A study of the relationship between academic achievement motivation and home environment among standard eight pupils. *Educational Research and Reviews, 5*(5), 213-217.

Murray, G. N., Low, J. B., Hollis, C., Cross, W. A., & Davis, M. S. (2007). Coordinated school health programs and academic achievement: A systematic review of the literature. *Journal of School Health, 77*, 589-600.

Myatt, M. (2013, Oct 20). *Data for dummies*. Retrieved from Supporting school improvement : http://marymyatt.com/blog/2013-10-20/data-for-dummies

Nes, F. v., Abma, T., Jonsson, H., & Deeg, D. (2010). Language differences in qualitative research: is meaning lost in translation? *European journal of ageing, 7*(4), 313–316.

Neuman, L. W. (2014). *Social research methods: Qualitative and quantitative approaches* (7th ed.). Essex: Pearson Education Ltd.

Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological review, 91*(3), 328-346.

Nicholls, J. G., Cobb, P., Wood, T., Yackel, E., & Patashnick, M. (1990). Assessing students' theories of success in mathematics: Individual and classroom differences. *Journal for Research in Mathematics education*, 109-122.

Noels, K. A., Pelletier, L. G., Clément, R., & Vallerand, R. J. (2000). Why are you learning a second language? Motivational orientations and self-determination theory. *Language learning, 50*(1), 57-85.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill Education.

Nyquist, J. B. (2003). *The benefits of reconstruing feedback as a larger system of formative assessment: A meta-analysis.* Nashville, TN: Vanderbilt University.

OECD. (2016). *EDUCATION IN CHINA: A Snapshot.*

Oga-Baldwin, W. Q., Nakata, Y., Parker, P., & Ryan, R. M. (2017). Motivating young language learners: A longitudinal model of self-determined motivation in elementary school foreign language classes. *Contemporary Educational Psychology, 49*, 140-150.

Ogunniyi, M. B. (1984). *Educational measurement and evaluation.* Nigeria: Longman.

Oxford, R., & Shearin, J. (1994). Language learning motivation: Expanding the theoretical framework. *The modern language journal , 78*(1), 12-28.

Öztürk, E. Ö. (2012). Contemporary motivation theories in educational psychology and language learning: An overview. *The International Journal of Social Sciences, 3*(1), 33-46.

Pae, T.-I. (2008). Second Language Orientation and Self-Determination Theory A Structural Analysis of the Factors Affecting Second Language Achievement. *Journal of Language and Social Psychology, 27*(1), 5-27.

Pajares, F. (2003). Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. *Reading &Writing Quarterly, 19*(2), 139-158.

Pallant, J. (2013). *SPSS survival manual.* UK: McGraw-Hill Education.

Pekrun, R. (1993). Facets of adolescents' academic motivation: A longitudinal expectancy-value approach. In P. Pintrich, & M. L. Maehr (Eds.), *Advances in motivation and achievement* (Vol. 8, pp. 139–189). Greenwich, CT: JAI.

Pekrun, R. (2009). Emotions at school. In K. R. Wentzel, & A. Wigfield (Eds.), *Handbook of motivation in school* (pp. 575–605). New York: Routledge.

Peng, Q. (2017). Formative Assessment in China and Finland : Teachers' perception, attitude and implementation.

Pham, T. H., & Renshaw, P. (2015). Formative assessment in Confucian heritage culture classrooms: activity theory analysis of tensions, contradictions and hybrid practices. *Assessment & Evaluation in Higher Education, 40*(1), 45-59.

Phillipson, S., & Phillipson, S. N. (2007). Academic expectations, belief of ability, and involvement by parents as predictors of child achievement: A cross-cultural comparison. *Educational Psychology, 27*, 329–348.

Pinter, A. (2007). Some benefits of peer–peer interaction: 10-year-old children practising with a communication task. *Language Teaching Research, 11*(2), 189-207.

Pintrich, P. R. (2000). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary educational psychology, 25*(1), 92-104.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology, 82*(1), 33-40.

Pintrich, P. R., & Schunk, D. H. (1996). *Motivation in education: theory, research, and applications.* Englewood Cliffs, NJ: Prentice Hall.

Pintrich, P. R., Smitch, D., Garcia, T., & McKeachie, W. J. (1991). A manual for the use of the Motivated Strategies for Learning Questionnaire (Technical Report 91-B-004). *The Regents of The University of Michigan*.

Ponce, H. R., Mayer, R. E., Figueroa, V. A., & López, M. J. (2018). Interactive highlighting for just-in-time formative assessment during whole-class instruction: effects on vocabulary learning and reading comprehension. *Interactive Learning Environments, 26*(1), 42-60.

Poole, A. (2016). Complex teaching realities' and 'deep rooted cultural traditions': Barriers to the implementation and internalisation of formative assessment in China. *Cogent Education*, 1-14.

Pop, E. L., Negru-Subtirica, O., Crocetti, E., Opre, A., & Meeus, W. (2016). On the interplay between academic achievement and educational identity: A longitudinal. *Journal of Adolescence, 47*, 135-144.

Qian, M., & Clark, K. R. (2016). Game-based Learning and 21st century skills: A review of recent research. *Computers in Human Behavior, 63*, 50-58.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science, 28*, 4-13.

Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing, 18*(4), 429-462.

Rea-Dickins, P., & Gardner, S. (2000). Snares and silver bullets: disentangling the constructt of formative assessment. *Language Testing, 17*(2), 215-243.

Rimfeld, K., Kovas, Y., Dale, P. S., & Plomin, R. (2016). True Grit and Genetics: Predicting Academic Achievement From Personality. *Journal of Personality and Social Psychology, 11*(5), 780-789.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417-458.

Robson, C., & McCartan, K. (2016). *Real world research: A resource for users of social research methods in applied settings* (4 ed.). London: Wiley.

Rochoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 247-252.

Rowe, M. B. (1974). Wait time and rewards as instructional variables, their influence on language, logic and fate control. *Journal of Research in Science Teaching, 11*, 81-94.

Rowntree, D. (1977). *Assessing Students: How shall we know them? .* London: Kogan Page.

Ryan, R. M., & Deci, E. L. (2009). Promoting self-determined school engagement: Motivation, learning, and well-being. In K. R. Wentzel, & A. Wigfield (Eds.), *Handbook of motivation in school.* New York: Routledge.

Sauders, M., Lewis, P., & Thornhill, A. (2000). *Research methods for business students* (2 ed.). London: Prentice-Hall.

Schunk, D. H. (1990). Goal setting and self-efficacy during self-regulated learning. *Educational psychologist, 25*(1), 71-86.

Schunk, D. H., & Pajares, F. (2004). Self-efficacy in education revisited: Empirical and applied evidence. In D. M. McInerney, & S. V. Etten (Eds.), *Big theories revisited* (pp. 115-138). Greenwich, CT: Information Age.

Schunk, D. H., & Pajares, F. (2009). Self-efficacy theory. In *Handbook of motivation at school* (pp. 35-53). New York: Routledge.

Scriven, M. (1967). *The Methodology of Evaluation.* Washington DC: American Education Research Association.

Seabi, J., & Payne, J. (2013). Effects of identity processing styles on academic achievemnt of first year university students. *International Journal of Educational Management, 27*(3), 311-322.

Sheard, M. K., Chambers, B., & Elliott, L. (2012). *Effects of technology-enhanced formative assessment on achievement in primary grammar.* Institute for Effective Education.

Shenkar, O., & Ronen, S. (1987). The Cultural Context of Negotiations: The Implications of Chinese Interpersonal Norms. *The Journal of Applied Behavioral Science, 23*(2), 263-275.

Shklarov, S. (2007). Double Vision Uncertainty: The Bilingual Researcher and the Ethics of Cross-Language Research. *Qualitative Health Research, 17*(4), 529-538.

Shumba, T., & Chisaka, C. B. (2013). Doing Research With Children. In M. T. Mukabeta, S. Hamandishe, & C. Nzombe (Eds.), *ACTION RESEARCH: Some Practical Ideas For Educational Practice* (pp. 50-53). Harare, Zimbabwe: Save the Children.

Smith, J. F., & Skrbiš, Z. (2017). A social inequality of motivation? The relationship between beliefs about academic success and young people's educational attainment. *British Educational Research Journal, 43*(3), 441-465.

Spinath, B. (2012). Academic achievement. In V. S. Ramachandran (Ed.), *Encyclopedia of human behavior* (2 ed., pp. 1–8). San Diego, CA: Academic Press.

Squires, A. (2009). Methodological challenges in cross-cultural qual- itative research: a research review. *International journal of nursing studies, 46*(2), 277-287.

SSAT. (2017). *Embedding Formative Assessment: schools' conclusions from a two-year project.* London.

Steinmayr, R., Meißner, A., Weidinger, A. F., & Wirthwein, L. (2017, Apr 28). *Academic Achievement*. Retrieved from Oxford Bibliographies: http://www.oxfordbibliographies.com/view/document/obo-9780199756810/obo-9780199756810-0108.xml

Stiggins, R. (2002). Assessment crisis: The Absence of Assessment for Learning. *The Phi Delta Kappan, 83*(10), 758-765.

Stiggins, R., & Chappuis, J. (2006). What a difference a word makes: Assessment for learning rather than assessment of learning helps students succeed. *Journal of Staff Development, 27*(1), 10-15.

Sung, Y. T., Chang, K. E., Chang, T. H., & Yu, W. C. (2010). How many heads are better than one? The reliability and validity of teenagers' self-and peer assessments. *Journal of Adolescence, 33*(1), 135-145.

Takagishi, H., Fujii, T., Koizumi, M., Schug, J., Nakamura, F., & Kameshima, S. (2015). The development of the effect of peer monitoring on generosity differs among elementary school-age boys and girls. *Frontiers in psychology, 6*(895).

Tan, C. (2017). Constructivism and pedagogical reform in China: Issues and challenges. *Globalisation, Societies and Education, 15*(2), 238-247.

Taras, M. (2010). Assessment for learning: assessing the theory and evidence. *Procedia-Social and Behavioral Sciences, 2*(2), 3015-3022.

Tell, C. A., Bodone, F. M., & Addie, K. L. (2000). *A Framework of Teacher Knowledge and Skills Necessary in a Standards-Based System: Lessons from High School and University Faculty.* New Orleans: Annual Meeting of the American Educational Research Association.

The State Council Information Office of the People's Repubic of China. (2016, Sept 14). *Report on Human Rights*. Retrieved from http://www.scio.gov.cn/zhzc/8/4/document/1491185/1491185.htm

Thomas, W. P., & Collier, V. P. (2002). *A National Study of School Effecitveness for Language Minority Students' Long-term Academic Achievement.* Center for Research on Education, Diversity and Excellence, Santa Cruz, CA.

Ting-Toomey, S. (2017). Conflict Face-negotiation Theory. In *Conflict Management and Intercultural Communication: The Art of Intercultural Harmony* (pp. 123-143).

Torrance, H., & Pryor, J. (1998). *Investigating Formative Assessment: Teaching, Learning and Assessment in the Classroom.* Maidenhead, Philadelphia: Open UP.

Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal, 27*(5), 615-631.

Tucker-Drob, E. M., & Briley, D. A. (2016). Genetically-Mediated Associations Between Measures of Childhood Character and Academic Achievement. *Journal of Personality and Social Psychology, 111*(5), 790-815.

Tunstall, P., & Gipps, C. (1996a). "How does your teacher help you to make your work better?" Children's understanding of formative assessment. *The Curriculum Journal, 7*(2), 185-203.

Tunstall, P., & Gipps, C. (1996b). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal, 22*(4), 389-404.

Turcker-Drob, E. M., & Briely, D. A. (n.d.). Gentically-Mediated Associations Between Measures of Childhood Charac.. .

Tzuriel, D., & Shamir, A. (2002). The effects of mediation in computer assisted dynamic assessment. *Journal of Computer Assisted Learning, 18*, 21-32.

Vallerand, R. J. (1997). Toward a hierarchical model of intrinsic and extrinsic motivation. *Advances in experimental social psychology, 29*, 271-360.

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briere, N. M., Senecal, C., & Vallieres, E. F. (1992). The Academic Motivation Scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educational and psychological measurement, 52*(4), 1003-1017.

Vazou, S., Gavrilou, P., Mamalaki, E., Papanastasiou, A., & Sioumala, N. (2012). Does integrating physical activity in the elementary school classroom influence academic motivation? *International Journal of Sport and Exercise Psychology, 10*(4), 251-263.

Verboom, G., Sijtsema, J., Verhulst, F., Penninx, B. W., & Ormel, J. (2014). Longitudinal associations between depressive problems, academic performance, and social functioning in adolescent boys and girls. *Developmental Psychology, 50*, 247e257.

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*, 1174–1204.

Wang, W. (2014). Students' perceptions of rubric-referenced peer feedback on EFL writing: A longitudinal inquiry. *Assessing Writing, 19*, 80-96.

Webb, E. J., Campbell, D. T., Schuwartz, R. D., & Sechrest, L. (2000). *Unobtrusive Measures: Nonreactive Research in the Social Siences* (revised ed.). Thousand Oaks, CA: SAGE.

Weiner, B. (1972). Attribution theory, achievement motivation, and the educational process. *Review of educational research, 42*(2), 203-215.

Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of educational psychology, 71*(1), 3-25.

Weiner, B. (1986). *An attributional theory of motivation and emotion.* New York: Springer-Verlag.

Weiner, B. (1992). *Human motivation: Metaphors, theories and research.* Newbury Park, CA: SAGE.

Weiner, G., Arnot, M., & David, M. (1997). Is the future female? Female success, male disadvantage, and changing gender patterns in education. In A. H. Halsey, P. Brown, & H. Lauder, *Education, Economy, Culture and Society.* UK: Oxford University Press.

Wellington, J., & Szczerbinski, M. (2007). *Research methods for the social sciences.* London: Continuum International Publishing Group.

Wentzel, K. R., & Wigfield, A. (2009). *Handbook of Motivation at School.* New York: Routledge.

Whetstone, S. (2011, Apr 14). *School Governing blog*. Retrieved from http://schoolgoverning.blogspot.co.uk/2011/04/understanding-attainment-achievement.html

Wigfield, A., & Cambria, J. (2010). Expectancy-value theory: Retrospective and prospective. In *The decade ahead: Theoretical perspectives on motivation and achievement* (pp. 35-70). Emerald Group Publishing Limited.

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary educational psychology, 25*(1), 68-81.

Wigfield, A., Eccles, J. S., & Rodriguez, D. (1998). The development of children's motivation in school contexts. In A. Iran-Nejad, & P. D. Pearson (Eds.), *Review of research in education* (Vol. 23, pp. 73–118). Washington, DC: American Educational Research Association.

Wigfield, A., Tonks, S., & Klauda, S. L. (2009). Expectancy-value theory. In K. R. Wentzel, & A. Wigfield (Eds.), *Handbook of Motivation at School* (pp. 55-74). New York: Routledge.

Wiliam, D. (2000). Formative assessment in mathematics part 3: The Learner's Role. *Equals: Mathematics and Special Education Needs, 6*(1), 19-22.

Wiliam, D. (2006). Formative Assessment: Getting the focus right. *Educational Assessment, 11*(3&4), 283-289.

Wiliam, D. (2007a). Keeping learning on track: Formative assessment and the regulation of learning. In F. K. Lester (Ed.), *Second handbook of mathematics teaching and learning* (pp. 1053–1098). Greenwich, CT: Information Age Publishing.

Wiliam, D. (2007b). Content then process: Teacher learning communities in the service of formative assessment. In D. Reeves (Ed.), *Ahead of the curve: the power of assessment to transform teaching and learning* (pp. 182-204). Bloomington, Indiana: Solution Tree.

Wiliam, D. (2011). *Embedded Formative Assessment.* Bloomington: Solution Tree.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*, 3-14.

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

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teacher developing assessment for learning: impact on student achievement. *Assessment in Education, 11*(1), 49-65.

Williams, M., & Burden, R. (1997). *Psychology for language teachers.* Cambridge: Cambridge University Press.

Wininger, S. R., & Norman, A. D. (2005). "Teacher-candidates" exposure to formative assessment in educational psychology textbooks: A content analysis. *Educational Assessment, 10*(1), 19-37.

Wong, K. -C., Lam, Y. R., & Ho, L. -M. (2002). The effects of schooling on gender differences. *British Educational Research Journal, 28*, 827–843.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson tests of achievement.* Itasca, IL: Riverside Publishing.

Wylie, E. C., & Lyon, C. J. (2015). The fidelity of formative assessment implementation: issues of breadth and quality. *Assessment in Education: Principles, Policy & Practice, 22*(1), 140-160.

Xiao, Y., & Carless, D. R. (2013). Illustrating students' perceptions of English language assessment: Voices from China. *RELC Journal, 44*(3), 319-340.

Xu, Y., & Carless, D. (2016). 'Only true friends could be cruelly honest': cognitive scaffolding and social-affective support in teacher feedback literacy. *Assessment & Evaluation in Higher Education, 42*(7), 1082-1094.

Yan, C. (2015). 'We can't change much unless the exams change': Teachers' dilemmas in the curriculum reform in China. *Improving Schools, 18*(1), 5-19.

Yan, J. X., & Horwitz, E. K. (2008). Learners' perceptions of how anxiety interacts with personal and instructional factors to influence their achievement in English: A qualitative analysis of EFL learners in China. *Language Learning, 58*(1), 151-183.

Yan, Z., & Cheng, E. C. (2015). Primary teachers' attitudes, intentions and practices regarding formative assessment. *Teaching and Teacher Education, 45*, 128-136.

Yu, G., & Jin, Y. (2014). English language assessment in China: Policies, practices and impacts. *Assessment in Education: Principles, Policy & Practice, 21*(3), 245-250.

Zach, S., Shoval, E., & Lidor, R. (2017). Physical education and academic achievement-literature review 1997-2015. *Journal of Curriculum Studies, 49*(5), 703-721.

Zhang, J. (2013). The utility of formative assessment in Chinese primary schools (Jiangsu Province). *Unpublished MA dissertation*.

Zhao, H. (2014). Investigating teacher-supported peer assessment for EFL writing. *ELT journal, 68*(2), 155-168.

Zhu, W. (2003). The one child family policy. *Archives of Disease in Childhood, 88*(6), 463-464.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational psychologist, 25*(1), 3-17.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary educational psychology, 25*(1), 82-91.

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American educational research journal, 29*(3), 663-676.

Zsolnai, A. (2002). Relationship between children's social competence, learning motivation and school achievement. *Educational psychology, 22*(3), 317-329.

Appendices

Appendix 1 A descriptive and analytic framework of the processes of formative classroom assessment

	Description	Possible teacher intentions	Possible positive effect for student
A	T communicates task criteria (what has to be done in order to complete the task) or negotiates them with S.	Communicating goals and success criteria; ensuring work is on target; adjusting pace of work.	Understanding of task and principles behind it.
B	T communicates quality criteria (what has to be done to do the task well) or negotiates them with S.	Enhancement of quality of future work; promotion of greater S independence.	Understanding of notions of quality to aid future self-monitoring.
C	T observes S at work (process).	Gain in understanding of why/how student has approached or achieved task.	Enhanced motivation due to T's attention.
D	T examines work done (product).	Gain in understanding of what S has done.	Enhanced motivation due to T's attention.
E	T asks principled question (seeks to elicit evidence of what S knows, understand or can do: substance). S responds.	Insight into S's knowledge, understanding or skills.	Rehearsal of knowledge, understanding, or skills; articulation of understanding to realise understanding.
F	T asks for clarification about what has been done, is being done or will be done: process; S relies.	Gain in understanding of what S has done and of S's understanding of the task.	Re-articulation of understanding; enhanced self-awareness and skills of summary, reflection, prediction, speculation.
G	T questions S about how and why specific action has been taken (meta-process and metacognitive questioning). S responds.	Gain in understanding of why/how the student has approached or achieved task. Promotion of deepened understanding and 'handover'.	Articulation of thinking-about-thinking; deepened understanding and 'handover'.
H	T critiques a particular aspect of the work or invites S to do so.	Enhancement of quality of future work; promotion of greater independence.	Articulating and interrogating quality criteria; enhanced understanding of quality issues; practice in self-monitoring.
I	T supplies information, corrects, or makes a	Communication of alternative or more	Enhancement of knowledge and/or understanding.

	counter-suggestion.	acceptable product.	
J	T gives and/or discusses evaluative feedback on work done with respect to: task, and/or effort and/or aptitude/capability (possibly with reference to past or future achievement: ipsative).	Influence on S's attributions and therefore motivation of S for further work.	Enhanced motivation and self-worth when realised in a context of empowerment; development of learning goals.
K	T suggests or negotiates with S what to do next.	Insight into ways forward for immediate further teaching of individual; refocusing S on curricular goals.	Insight in ways to continue working and learning. Deepening of understanding of process/principle.
L	T suggests or negotiates with S what to do next time.	Insights into ways forward for planning of group activities.	Deepening of understanding of principle/process.
M	T assigns mark, grade or summary judgement on the quality of this piece of work or negotiates an agreed one with S.	Information for SA; communication of quality criteria; teaching/modelling skills of assessment for self-assessment.	Information about present achievement with respect to longer term goals.
N	T rewards or punishes student, or demonstrates approval/disapproval.	Improvement or maintenance of relationship with students; enhancement of motivation.	

Appendix 2 Questionnaire for students (Main study original-Chinese)

你好，我叫张家维，是谢菲尔德大学教育学院的一名博士生。这是一份调查问卷，一共有三部分：第一部分是在英语学习中的动机，第二部分是有关形成性评价在小学英语教学中的现状，第三部分是你对形成性评价的看法 （形成性评价：在本研究中指以提高英语学习为目的的各种评价手段）。这份问卷可能会占用你 10 分钟的时间。答案没有对错，也不会被除我之外的人知道。而且数据仅用于博士论文的撰写，以及日后学术会议的展示。你的个人信息将会被匿名编码，不会在任何刊物中被认出，所以请放心填写你觉得最符合你的答案。

调查问卷

在以下的问题中，请你根据实际情况，选择你对以下陈述的同意程度。例如：

很不同意= 1; 不同意 = 2; 不知道 = 3; 同意 = 4; 非常同意 = 5

英语对我来说很重要	1	2	3	4	5

如果你觉得英语对你的确很重要，你对此非常同意，那么就请在 5 (非常同意) 上画圈，像ʘ一样。

以下是问卷的主要内容，请选择你心中的答案：

情形陈述	很不同意	不同意	不确定	同意	非常同意
对我来说英语很重要	1	2	3	4	5
对我来说英语很有意思	1	2	3	4	5
对我来说英语很有用	1	2	3	4	5
我愿意在学习英语上花时间	1	2	3	4	5
对我来说老师说的学习目标很容易	1	2	3	4	5
对我来说老师说的学习目标很明确	1	2	3	4	5
我学英语是因为我想比别人更优秀	1	2	3	4	5
我学英语是因为我想掌握这门语言	1	2	3	4	5
我觉得我有能力把英语学好	1	2	3	4	5
我觉得我学英语很努力	1	2	3	4	5
我对我的英语有很高的信念	1	2	3	4	5
我对我的英语有很强的信心	1	2	3	4	5
成功学习英语是外界需要我获得的	1	2	3	4	5
成功学习英语是我自己所想要的	1	2	3	4	5
我喜欢学习英语时成功的感觉	1	2	3	4	5
我害怕学习英语时失败的感觉	1	2	3	4	5
我的老师对我学英语有很大影响力	1	2	3	4	5
我的家人对我学英语有很大影响力	1	2	3	4	5
我的朋友对我学英语有很大影响力	1	2	3	4	5
我学英语是因为我喜欢西方文化	1	2	3	4	5

形成性评价在英语教学中的现状

情形陈述	从来不	偶尔	一般	经常	一直是
老师在英语课上会和告诉我们这堂课的要上的内容	1	2	3	4	5
老师上课会问谁/哪儿/何时/什么等开放式问题	1	2	3	4	5
老师在课堂上会问用 yes/no 就可以回答的问题	1	2	3	4	5
老师会用肯定性的词汇确认我的答案（比如 yes, good）	1	2	3	4	5
当我回答问题遇到困难时候，老师会提供我一些引导	1	2	3	4	5
当我回答问题不对的时候，老师会直接告诉我答案	1	2	3	4	5
当我完成任务后，我的同学会鼓励我，或者给我建议	1	2	3	4	5
当我回答问题不对的时候，我的同学会做出负面的反应	1	2	3	4	5
我在课堂学习中会做一做自己评价自己的事情	1	2	3	4	5

你对形成性评价的看法

情形陈述	很不同意	不同意	不确定	同意	非常同意
知道这堂课要学什么　　让我更想学习英语	1	2	3	4	5
老师问谁/什么/哪儿/何时等问题让我更想学习英语	1	2	3	4	5
用 yes/no 回答的问题　　让我更想学习英语	1	2	3	4	5
老师肯定鼓励我的回答　　让我更想学习英语	1	2	3	4	5
说错的时候老师引导我　　让我更想学习英语	1	2	3	4	5
说错时老师直接纠正我　　让我更想学习英语	1	2	3	4	5
同学的鼓励肯定　　让我更想学习英语	1	2	3	4	5
同学指出我的缺点　　让我更想学习英语	1	2	3	4	5
自己对自己做出反思评价　　让我更想学习英语	1	2	3	4	5

个人信息

班级	四（　　）班	学号	号

非常感谢您的参加！如果您需要关于本研究的后续信息，请用以下电话号码或者邮箱地址联系我：

jiawei.zhang@sheffield.ac.uk
(+44)07851484329 (英国)
(+86)18344762989 (中国)

Appendix 2.1 Questionnaire for students (Main study translated-English)

Good morning! My name is Jiawei Zhang; I am a PhD student in the School of Education, at The University of Sheffield. This is a questionnaire which contains two parts: the motivation issues in English learning process and your perception on Formative Assessment. (Formative Assessment in this study is different assessment strategies which can promote your English learning.) I wonder if you could take just 10 minutes to complete this questionnaire. Your answers are strictly confidential and will only be used for statistical purposes for my PhD thesis; your teacher and parents will not know what you put on this. Please feel free to choose your answers.

Questionnaire of Motivation in English learning and Perceptions on Formative Assessment

In the following questions, you are being asked to indicate your level of agreement or disagreement with each statement by indicating whether you strongly disagree (SD), disagree (D), are undecided (U), agree (A), or strongly agree (SA). Here is an example:

Strongly Disagree = 1; Disagree = 2; Undecided = 3; Agree = 4; Strongly Agree = 5					
English is important to me.	1	2	3	4	5

If you feel that English **is** important to you, you strongly agree with the statement, and then you may circle number 5, like: ⑤, indicating 'Strongly Agree'.

Here is the main questionnaire, please choose YOUR answers.

Statements	SD	D	U	A	SA
English is important to me	1	2	3	4	5
English is interesting to me	1	2	3	4	5
English is useful to me	1	2	3	4	5
I am willing to spend time on learning English	1	2	3	4	5
Learning aims are easy to me	1	2	3	4	5
Learning aims are specific to me	1	2	3	4	5
I learn English because I want to be better than others	1	2	3	4	5
I learn English because I want to master it	1	2	3	4	5
I have the ability to learn English well	1	2	3	4	5
I work hard in learning English	1	2	3	4	5
I have high faith in my English	1	2	3	4	5
I am strongly confident in my English	1	2	3	4	5
I need to succeed in English learning	1	2	3	4	5
I want to succeed in English learning	1	2	3	4	5
I like the feeling of being successful in English learning	1	2	3	4	5
I am afraid of failure in English learning	1	2	3	4	5
I learn English is because of my English teacher	1	2	3	4	5
I learn English is because of my parents	1	2	3	4	5
I learn English is because of friends	1	2	3	4	5
I learn English because of my school requests	1	2	3	4	5
I learn English because I like western culture	1	2	3	4	5

Current situation of Formative Assessment in teaching

Statements	SD	D	U	A	SA
Teacher shares learning goals with us every class.	1	2	3	4	5
Teacher asks a lot of wh- questions.	1	2	3	4	5
Teacher asks a lot of yes/no questions.	1	2	3	4	5
Teacher always confirm my answers, such as 'yes, good'.	1	2	3	4	5
Teacher always gives me suggestions when I struggle with the question.	1	2	3	4	5
Teacher always correct me when I answer the question wrong.	1	2	3	4	5
My classmates always encourage me after I complete a task.	1	2	3	4	5
My classmates always behave negatively, if I fail to answer the question.	1	2	3	4	5
I always complete the self-assessment form at the end of a lesson.	1	2	3	4	5

Your Perception of Formative Assessment

Statements	SD	D	U	A	SA
I find the goals help me learn English better.	1	2	3	4	5
Wh- questions make me want to learn English more.	1	2	3	4	5
Yes/no questions make me want to learn English more.	1	2	3	4	5
I find affirmative assessment is very useful to my English learning.	1	2	3	4	5
I find suggestive assessment is very useful to my English learning.	1	2	3	4	5
I find corrective assessment is very useful to my English learning.	1	2	3	4	5
The positive feedback from my classmates can motivate me.	1	2	3	4	5
The negative feedback from my classmates can motivate me.	1	2	3	4	5
Self-assessment form is useful to my English learning.	1	2	3	4	5

Personal Information

School:			
Class:		Registration number:	
Age:		Gender:	M □ F □

The End.

Thank you very much for participating! If you need more information about the following research, please contact me with the email address or telephone numbers below:

jiawei.zhang@sheffield.ac.uk
(+44)07851484329 (UK)
(+86)15358260308 (China)

Appendix 3 Questionnaire for students (Pilot study translated-English)

Good morning! My name is Jiawei Zhang; I am a PhD student in the School of Education, at The University of Sheffield. This is a questionnaire which contains two parts: the motivation issues in English learning process and your perception on Formative Assessment. (Formative Assessment in this study is different assessment strategies which can promote your English learning.) I wonder if you could take just 10 minutes to complete this questionnaire. Your answers are strictly confidential and will only be used for statistical purposes for my PhD thesis; your teacher and parents will not know what you put on this. Please feel free to choose your answers.

Questionnaire of Motivation in English learning and Perceptions on Formative Assessment

In the following questions, you are being asked to indicate your level of agreement or disagreement with each statement by indicating whether you strongly disagree (SD), disagree (D), are undecided (U), agree (A), or strongly agree (SA). Here is an example:

Strongly Disagree = 1; Disagree = 2; Undecided = 3; Agree = 4; Strongly Agree = 5					
English is important to me.	1	2	3	4	5

If you feel that English **is** important to you, you strongly agree with the statement, and then you may circle number 5, like: ⑤, which indicate 'Strongly Agree'.

Here is the main questionnaire, please choose YOUR answers.

Statements	SD	D	U	A	SA
English is important to me	1	2	3	4	5
English is interesting to me	1	2	3	4	5
English is useful to me	1	2	3	4	5
I am willing to spend time on learning English	1	2	3	4	5
Learning aims are easy to me	1	2	3	4	5
Learning aims are specific to me	1	2	3	4	5
I learn English because I want to be better than others	1	2	3	4	5
I learn English because I want to master it	1	2	3	4	5
I have the ability to learn English well	1	2	3	4	5
I work hard in learning English	1	2	3	4	5
I have high faith in my English	1	2	3	4	5
I am strongly confident in my English	1	2	3	4	5
I need to succeed in English learning	1	2	3	4	5
I want to succeed in English learning	1	2	3	4	5
I like the feeling of being successful in English learning	1	2	3	4	5
I am afraid of failure in English learning	1	2	3	4	5
I learn English is because of my English teacher	1	2	3	4	5
I learn English is because of my parents	1	2	3	4	5
I learn English is because of friends	1	2	3	4	5
I learn English because of my school requests	1	2	3	4	5
I learn English because I like western culture	1	2	3	4	5

Your Perception of Formative Assessment

Statements	SD	D	U	A	SA
Teacher shares learning goals with us every class.	1	2	3	4	5
I find the goals help me learn English better.	1	2	3	4	5
Teacher asks a lot of wh- questions.	1	2	3	4	5
Wh- questions make me want to learn English more.	1	2	3	4	5
Teacher asks a lot of yes/no questions.	1	2	3	4	5
Yes/no questions make me want to learn English more.	1	2	3	4	5
Teacher always confirm my answers, such as 'yes, good'.	1	2	3	4	5
I find affirmative assessment is very useful to my English learning.	1	2	3	4	5
Teacher always gives me suggestions when I struggle with the question.	1	2	3	4	5
I find suggestive assessment is very useful to my English learning.	1	2	3	4	5
Teacher always correct me when I answer the question wrong.	1	2	3	4	5
I find corrective assessment is very useful to my English learning.	1	2	3	4	5
My classmates always encourage me after I complete a task.	1	2	3	4	5
The positive feedback from my classmates can motivate me.	1	2	3	4	5
My classmates always behave negatively, such as laugh at me, if I fail to answer the question.	1	2	3	4	5
The negative feedback from my classmates can motivate me.	1	2	3	4	5
I always complete the self-assessment form at the end of a lesson.	1	2	3	4	5
Self-assessment form is useful to my English learning.	1	2	3	4	5

Personal Information

School:			
Class:		Registration number:	
Age:		Gender:	M □ F □

The End.

Thank you very much for your participation! If you need more information about the following research, please contact me with the email address or telephone numbers below:

jiawei.zhang@sheffield.ac.uk
(+44)07851484329 (UK)
(+86)15358260308 (China)

Appendix 4 English quiz sample (Main study original-Chinese)

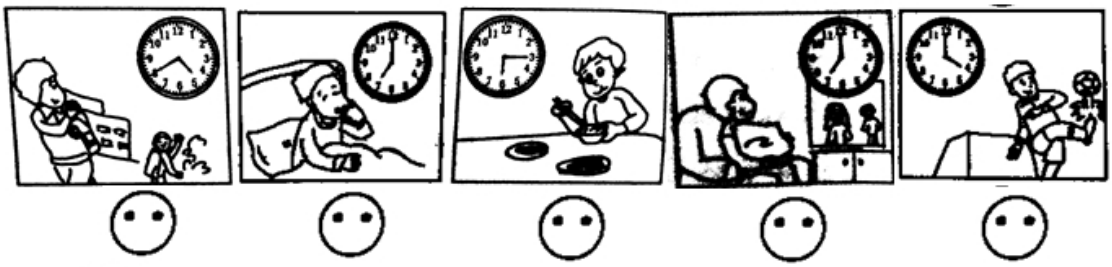
Quiz for Week 3

你的学号: _____ **你同桌**的学号: _____

请听单词，选出你所听到的单词

( )	1. A morning	B afternoon	C evening
( )	2. A usually	B Friday	C play
( )	3. A home	B school	C homework
( )	4. A football	B basketball	C tennis
( )	5. A breakfast	B lunch	C dinner

请根据听力，判断图片是否正确，用笑脸/哭脸回答



连词成句

school let's to go (.)

_____.

you home do when go (?)

_____?

完成句子

我晚上看电视。

I _____ _____ in the evening.

他们下午有体育课。

They have PE _____ _____ _____.

我们在十二点钟吃午饭。

We _____ _____ _____ twelve.

请朗读课文 Unit 3 Story Time，并根据 朗读 水平，互相评分（满分为 10 分，请用**整数**打分）

自己**读得**怎么样？_____分	我同桌**读得**怎么样？_____分

请根据这一周学习的内容和同桌一起编对话（至少每人 5 句话，包括问好和再见），并根据 对话 水平互相评分（满分为 10 分，请用**整数**打分）

自己**说得**怎么样？_____分	同桌**说得**怎么样？_____分

朗读与对话的评分标准（请对照标准评分）

1. 发音准确吗？	2. 认读正确吗？有多词、漏词吗？
3. 自然流畅、速度适中吗？有重复、断断续续吗？	4. 如果出现错误，能够自己纠正吗？
5. 带有恰当的情绪吗？	6. 语法正确吗？（仅适用于对话）

Appendix 4 English quiz sample (Main study translated-English)

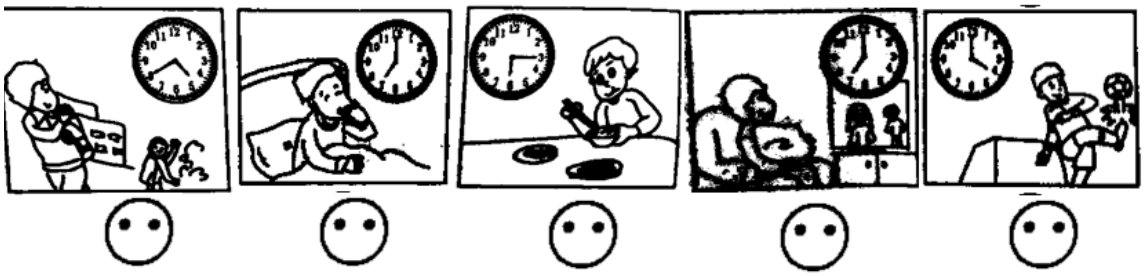
Quiz for Week 3

Your registration number: _____ **Your desk mate's registration number:** _____

Please listen to the recording, choose the right word

( )	1. A morning	B afternoon	C evening
( )	2. A usually	B Friday	C play
( )	3. A home	B school	C homework
( )	4. A football	B basketball	C tennis
( )	5. A breakfast	B lunch	C dinner

Please draw smile or cry face according to the recording



Please put all the words together make a proper sentence

school let's to go (.)

_____.

you home do when go (?)

_____?

Please complete the sentence

我晚上看电视。

I _____ _____ in the evening.

他们下午有体育课。

They have PE _____ _____ _____.

我们在十二点钟吃午饭。

We _____ _____ _____ twelve.

Please read Unit 3 Story Time，rate each other according to the criteria below (from 1 to 10)

How did I do in reading? _____	How did my desk mate do in reading? _____

Please make a conversation with your desk mate according to Unit 3, 5 sentences at least for each person, rate each other according to the criteria below (from 1 to 10)

How did I do in speaking? _____	How did my desk mate do in speaking? _____

Criteria for reading and speaking

1. Is the pronunciation correct?	2. Is there any extra or missing words?
3. Is it fluent, with moderate speed? Is it repeated or broken?	4. Can you/desk mate correct oneself if there is a mistake?
5. Is it in the suitable emotion?	6. Is it grammatically correct? (for speaking only)

Appendix 5 Interview guide for teachers (Main study original-English)

Interview guide for English teachers

Introduction

Thank you for taking part in the interview. First of all, I can assure you that you will remain completely anonymous, and your name will not be recorded on any document in this research.
Also I would like to ask you for the permission to record this interview. The main reason behind this recording is to ensure the accuracy of data. It will also make the analysis easier.
Do you have any questions? If you don't have any further question, I would like to introduce you the subject of this interview: Formative Assessment.
In this interview, I would like to ask questions about your understanding on the uses of the FA strategies and their impact on students' motivation and achievement.

Stage		Question	Follow-up questions
General	Q1	What Formative Assessment is to you?	Do you know any specific strategy of Formative Assessment?
	Q2	Do you use 'sharing learning aims, 'questioning', 'feedback', 'peer-assessment', and 'self-assessment' in your teaching?	The frequency and the quality of these strategies. Why do you think different teachers adopt different approaches? What impact do you see on the students' learning?
Specific	Q3	Let's talk about 'sharing learning aims. What do you think about it?	How/how often do you use it? Can this motivate students? Why? Can this help students with their achievement? How?
	Q4	Let's talk about 'questioning'. What do you think about it?	How/how often do you use it? Can this motivate students? Why? Can this help students with their achievement? How?
	Q5	Let's talk about 'feedback'. What do you think about it?	How/how often do you use it? Can this motivate students? Why? Can this help students with their achievement? How?
	Q6	Let's talk about 'peer-assessment'. What do you think about it?	How/how often do you use it? Can this motivate students? Why? Can this help students with their achievement? How?
	Q7	Let's talk about 'self-assessment'. What do you think about it?	How/how often do you use it? Can this motivate students? Why? Can this help students with their achievement? How?
	Q8	Do you think there are other Formative Assessment	How/how often do you use them? Can these strategies motivate

	strategies that you use in daily teaching?	students? Why? Can these strategies help students with their achievement? How?
Closure	We seem to have covered a great deal of ground and you have been very patient. But do you think there is anything we've missed out about FA and its impact on students' motivation and achievement? Please feel free to give me your suggestions. Do you have any other comments about what we have discussed, or about the research as a whole? Do you want to see a transcript of the interview? Thanks again for helping me with this research.	

Appendix 6 Interview guide for students (Main study original-Chinese)

首先感谢你们参加此次采访，你们的名字，照片都不会被播出去，拍视频的目的只是为了写文章的时候确保你们的答案的准确性。所以，请问可以继续拍摄嘛？那今天的采访是关于**形成性评价**和它对你学习动机和成就的影响。课堂上的形成性评价是来自于老师和同伴以及你自己的，可以帮助你学习英语的任何评价手段，包括语言性的和动作性的。在这次采访里，我们即将讨论一些有关这些评价对你们英语学习的动力以及成就有哪些影响的话题。

我们首先谈一谈 **分享学习目标**，分享学习目标就是老师告诉大家这节课需要学哪些内容，或者希望下课之前大家能完成哪些目标。
1.你的老师用这种教学手段吗？你觉得老师分享的目标简单吗？明确吗？
2.你喜欢这样的方式吗？为什么？
3.你觉得这样的方式，能激发你对英语学习的动力吗？
4.你觉得这样的方式，能提高你英语学习的成就吗？

下面我们谈一谈 **课堂提问**，课堂提问的时候，老师提的问题可能会归纳为两种，一种是开放式问题，一种是闭合式问题。开放式问题比如：…闭合式问题比如：…
1.老师在课堂中会经常问什么样的问题？请举个例子？
2.你比较喜欢哪一种问题？为什么？
3.你觉得还有其他不同类型的问题嘛？比如？
4.你觉得不同的提问方式，能激发你对英语学习的动力吗？
5.你觉得不同的提问方式，能提高你英语学习的成就吗？

下面我们谈一谈 **课堂反馈**，课堂反馈包括老师口头上的反馈和肢体语言的反馈。
当你回答完问题后，老师经常怎么样评价或者反馈你？比如说？
老师在课堂中还会在什么情况下给你评价和反馈？是什么样的评价？确定/建议/纠正
你喜欢什么样的评价或者反馈？
4.你觉得不同的反馈方式，能激发你对英语学习的动力吗？
5.你觉得不同的反馈方式，能提高你英语学习的成就吗？

下面我们谈一谈 **同伴评价**， 同伴评价是指你的同学在课堂上对你的包括正面和负面的，包括言语上或动作上的评价。
你的同伴在课堂上评价过你吗？请举个例子？
你在课堂上评价过你的同伴吗？请举个例子？
你觉得你喜欢同伴之间的评价吗？喜欢怎样的评价？
你觉得不同的同伴评价，能激发你对英语学习的动力嘛？
你觉得不同的同伴评价，能提高你英语学习的成就嘛？

下面我们谈一谈 **自我评价**，自我评价是指你自己在课堂上对自己的做出的任何评价反馈反思。
你在课堂上评价过自己的英语学习吗？怎么评价？
你喜欢自我评价这种方式嘛？为什么？

你觉得课堂上的自我评价，能激发你对英语学习的动力吗？
你觉得课堂上的自我评价，能提高你英语学习的成就吗？

除了我们刚刚谈到的几种评价，你觉得课堂上还有哪些来自于老师、同伴以及自己的，对英语学习有帮助的评价？你觉得它对你对英语学习的动机和成就有什么帮助嘛？

最后还想请问，你觉得这次采访里，有关于形成性评价以及英语学习动机的，还有哪些你想讲，但之前没有讲到的内容？你对这次采访整体还有什么疑问吗？你需要此次采访的文字稿吗？
谢谢参加本次采访！

Appendix 6.1 Interview guide for students (Main study translated-English)

Interview schedule for English learners

Interview length: _____ minutes
Date: _____ Time: _____

About the interviewee
Profession: English learner at _____ Primary School, Class_____
Age: _____ Gender: _____

Introduction

Thank you for being willing to take part in an interview in this study. Can I first of all assure you that you, if you wish so, will remain completely anonymous and no records of the interview will be kept with your name on them.

Also I would like to ask you for permission to audio record this interview. Then main reason behind this recording is to have the set of accurate data – your responses and opinions. Also it will facilitate the analysis of the data we have to conduct in next stage.

If you don't have any further question I would like briefly to introduce you to the subject of this interview.

As you can see we brought with us 5 different types of Formative Assessment strategies in English classes. Formative Assessment is the assessment which helps you with your learning, such as sharing learning goals, teacher asking questions, teacher giving you feedback, your classmates assessing you, and your own assessment. In this interview, I would like to ask questions about your understanding on the uses of the strategies and their impact on your motivation and achievement.

Stage		Question	Follow-up questions
Specific	Q1	Let's talk about "sharing learning aims". What do you think about it?	Does your teacher share learning aims with you? (What kind of aims?) Do you like this? / Can you accept it? Does this help you with your English learning?
	Q2	Let's talk about "questioning". What do you think about it?	What kind of questions do you like? Does this help you with your English learning?
	Q3	Let's talk about "feedback". What do you think about it?	How does your teacher give your feedback? (Please give me some examples.) Do you like them? / Which type of feedback you like the most? Does this help you with your English learning?
	Q4	Let's talk about "peer-assessment". What do you think about it?	Do your classmates assess you in English learning? (Please give me some examples.) Do you like this?

			Does this help you with your English learning?
	Q5	Let's talk about "self-assessment". What do you think about it?	Do you assess yourself in English learning? (Please give me some examples.) Do you like this? Does this help you with your English learning?
	Q6	Do you think there are other assessment actions that can motivate you or help you with your achievement?	
Closure	We seem to have covered a great deal of ground and you have been very patient. But do you think there is anything we've missed out? Please feel free to give me your suggestions. Do you have any other comments about what we have discussed, or about the research as a whole? Do you want to see a transcript of the interview? I will send you a summary of the research findings some time in 2017, probably in April or May. Thanks again for help me with this research study.		

Appendix 7 Ethical approval in Confirmation Review Form

Application 000328

Section A: Applicant details

Date application started:
Tue 20 May 2014 at 20:36

First name:
Jiawei

Last name:
Zhang

Email:
jiawei.zhang@sheffield.ac.uk

Programme name:
PhD

Module name:
Education FT
Last updated:
23/01/2015

Department:
School of Education

Applying as:
Postgraduate research

Research project title:
The Impact of Formative Assessment on Learners' Motivation and Achievement in Language Learning - A Study with English Learners in Primary Schools in China

Similar applications:
- not entered -

Section B: Basic information

Supervisor

Name	Email
Tim Herrick	t.herrick@sheffield.ac.uk

Proposed project duration

Start date (of data collection):
Sun 4 January 2015

Anticipated end date (of project)
Sun 30 April 2017

3: URMS number (where applicable)

URMS number
- not entered -

Suitability

Takes place outside UK?
Yes

Involves NHS?
No

Healthcare research?
No

ESRC funded?
No

Involves adults who lack the capacity to consent?
No

Led by another UK institution?
No

Involves human tissue?
No

Clinical trial?
No

Social care research?
No

Vulnerabilities

Involves potentially vulnerable participants?
Yes

Involves potentially highly sensitive topics?
No

Section C: Summary of research

1. Aims & Objectives

This research focuses the Formative Assessment which used in the primary school English class in China, as well as the impact of formative assessment on learners' motivation and English achievement. The aim is to explore the relation between formative assessment and motivation, formative assessment and achievement. Eventually, I hope the research may help the English learning and teaching in China.

2. Methodology

For quantitative data collection: I will use questionnaires with student participants in Week 2 in the pilot study and Week 1, Week 10 in the main study. The questionnaire is about motivation in learning English in China and students' perception on formative assessment that used in their English classes. Quizzes on every Friday and a few more tests (in Week 2 in the pilot study and Week 5, Week 10 in the main study) will be conducted in order to draw a picture of students' achievement.

For qualitative data collection: I will use interviews with both teacher and student participants. Because of the great number of learners, group interview will also be used with students. During the interview, we are going to talk about formative assessment, and learners' motivation and achievement, to give me a sense of the impact of formative assessment in English classes in China. After the pilot study, all the interview questions will be modified, and applied in the interviews in the main study.

3. Personal Safety

Raises personal safety issues? No

- not entered -

Section D: About the participants

1. Potential Participants

a) 2 English teachers from Chinese primary schools
They come from 2 Chinese primary schools. They teach English for 2 classes in each school.
b) 200-240 Chinese students who are learning English
They are the students of the teacher participants. Normally, a Chinese primary school class has 50-60 students. Therefore, for 4 classes, the number of students participants is between 200-240.

2. Recruiting Potential Participants

There are 15-20 primary schools in the urban area in the target city, about 7 classes in each grade (on average), and 50-60 students in each class. Also, English is a compulsory course for everyone in the school. Therefore, 5,250-8,400 10-year-old students are learning English in primary schools.
Two schools with similar reputation and teaching quality will be selected. The data is not identifiable, also, the nature of the research is about individual learning and development, which it would be inappropriate to overly-generalise. Choosing two school is my choice to support the project which is aimed at enlarging sample while not being too concentrated in a single school.
The headteachers will be contacted firstly. After gaining their approval for conducting research in their schools, I will get in touch with all English teachers who teach English to Grade 4 and 5 (about 10-year-old). We will have an English teacher group meeting in each school to discuss the details about the project. Information sheet will be sent during the meeting. Anyone who is interested in my research can get in touch by ringing me after the discussion. I will put candidates' names into a box. One English teacher from each school will be randomly chosen by drawing the names out of the box.
After choosing the teacher participants, I will then send information sheet and consent forms to their students. All students will be asked to bring the forms back home and get the signature from the parents or the guardians. On the basis of the feedback from consent forms, all student who give me the assent will become the student participants in my research.
At that point, both teacher and student participants are recruited.

2.1. Advertising methods

Will the study be advertised using the volunteer lists for staff or students maintained by CiCS? - *not entered* -

- *not entered* -

3. Consent

Will informed consent be obtained from the participants? (i.e. the proposed process) Yes

Information sheet and consent form for students will be translated into Chinese and be distributed to all participants. Students will bring the documents back home to their parents or guardians. Any questions are welcome. Some technical terms will be explained in daily Chinese, and some simple examples will be introduced.
Information sheet and consent form for teachers will be distributed and they are in English.
The participants can join in this study only if they hand the consent form in with a signature from themselves, their parents or their guardians. If any participants don't give the consent, they will still be asked to attend all the activities. But the data from them won't be analysed. All the questionnaires, quizzes, and tests will be signed by students only with their registration number. Therefore, the data can be recognised and sorted.

4. Payment

Will financial/in kind payments be offered to participants? No

5. Potential Harm to Participants

What is the potential for physical and/or psychological harm/distress to the participants?

1. More quizzes and exams for students.
2. More workload to teachers, e.g. mark the test paper.
3. Have one more teacher staying the classroom when having English lessons.

How will this be managed to ensure appropriate protection and well-being of the participants?

1. Design short but efficient quizzes and exams.
2. Inform the participants with the upcoming extra work, ensure every participant is prepared before they decide to join in.
3. Work as an intern teacher in the class (which happens very often in China).

Section E: About the data

1. Data Confidentiality Measures

Data will be stored on an external hard drive which will be encrypted for protection.

2. Data Storage

Who will have control of, and act as the custodian for, the data generated by the project?--The researcher: me.
Where the analysis of the data from the project will take place and who will analyse the data?--In the UK, with office computer, and the data will be analysed by me.
Whether any encryption or other anonymisation will be used and at what stage?--All the name of schools, classes and participants will be anonymous.
Who will have access to the data generated by the project?--Me and my supervisor, only during the analysis process.
Whether the data will be available for use in future research projects--Yes
When (if ever) the data will be destroyed-- After the submission of final thesis.
Whether your research is externally funded and whether it has met the requirements of the funder with regards to data storage and management--No funding

The hardcopy of questionnaire will be destroyed right after inputting all the data into SPSS software package.
Interview audio files will be stored in PC with encryption, and transcribed both in office and at home. The audio files will be destroyed right after the full transcript is done.
All the data will be deleted or destroyed after the submission of final thesis.

Section F: Supporting documentation

Information & Consent

Participant information sheets relevant to project?
Yes

Document 004622 (Version 1)	All versions
Document 004623 (Version 1)	All versions
Document 004624 (Version 1)	All versions
Document 004625 (Version 1)	All versions
Document 004626 (Version 1)	All versions
Document 004627 (Version 1)	All versions

Consent forms relevant to project?
Yes

Document 004567 (Version 1)	All versions
Document 004568 (Version 1)	All versions
Document 004569 (Version 1)	All versions

Additional Documentation

Document 004570 (Version 1)	All versions
Document 004631 (Version 1)	All versions
Document 004744 (Version 1)	All versions
Document 004745 (Version 1)	All versions
Document 004746 (Version 1)	All versions
Document 004747 (Version 1)	All versions
Document 004748 (Version 1)	All versions

Document 004749 (Version 1) All versions

External Documentation

- not entered -

Section G: Declaration

Signed by:
Jiawei Zhang
Date signed:
Wed 21 January 2015 at 17:37

Offical notes

- not entered -

Appendix 8 Information sheet for students (Main study original-Chinese)

给学生的一封信

亲爱的同学，你好：

欢迎你参加本次调研。此研究项目仅作英国谢菲尔德大学教育学博士学位论文撰写之用途。在你决定是否继续参加之前，请和家长一起仔细阅读并讨论以下内容。此次调研内容是在中国小学生学习英语过程中，形成性评价（形成性评价是指任何为了提升学习的评价，本研究中具体表现为五大技巧：分享学习目标，提问，反馈，同学互评，自我评价）对学生动机和成就的影响；目的是为了了解形成性评价和学生动机以及成就的关系，为中国的英语教学提供帮助。本次研究分为两个部分：初步研究（两周）和主要调查（十周）。这份文件是针对**主要调查**的一封信息信。

你并没有一定要参加此次研究，你可以在任何时间段退出。如果你决定要参加，在此期间，你将被要求参加一系列有关于形成性评价/动机和成就的活动，例如：

照常上英语课；

在主要调查**第一/十周**完成两份问卷；

在主要调查**第五/十周**完成两次测验；

在每周结束时完成一份随堂检测；

可能参加单独采访或者小组采访。

所有的采访将会被录音，但录音和其他数据都只会被用于博士论文的撰写，或者学术会议的展示。**由于整个过程都通过匿名进行，所以你的信息不会在任何公开出版物中被认出。**

参加本次调研，有可能使你的英语学得更加扎实，并且让你认识到形成性评价在英语学习中的作用。但是，与此同时，你可能对问卷、测试，以及每周的随堂检测感到不舒服。

如果由于不可抗因素，本次调研提前被中止，所有收集到的数据会被销毁。再次继续调研时，之前的调查对象和数据都不会被录用，新的调查对象将通过和之前一样的办法被选出。

此次调查已经通过谢菲尔德大学教育学院伦理委员会的审查。如果在过程中，你发现有不当的地方，可以随时和我联系，或者退出此次调查。你也可以向我的指导老师：Tim Herrick/学院领导反映。如果他们的反馈不能让你满意，你还可以联系谢菲尔德大学秘书处。

如果你对调查结果感兴趣，可以联系我获取研究结果的复印件。如果还有其他问题，欢迎你打我在中国的电话：18344762989；或者发邮件给我。

非常感谢你的帮助！

研究者：张家维：Jiawei.zhang@sheffield.ac.uk
学术导师：Tim Herrick：t.herrick@sheffield.ac.uk
秘书处：registrar@sheffield.ac.uk

Appendix 8.1 Information sheet for students (Main study translated-English)

A letter to students

Dear student, hello,

Thank you for attending this research, this research is being undertaken for a Doctor of Philosophy degree at the University of Sheffield in the UK.

Before you decide to continue, please read the following information and discuss the details with your parents or guardians carefully. This research studies the impact of formative assessment (formative assessment is any assessment that used for improving learning, specifically defined as five main techniques in current study: sharing learning goals, questioning, feedback, peer-assessment and self-assessment) on students' motivation and achievement during the process of English learning; the aim of the research is to investigate the relation between formative assessment and motivation/achievement, and hopefully, to provide some help to English teaching and learning in China. This research is divided into two: a pilot study (2 weeks) and a main study (10 weeks). This letter is about the main study.

You don't have to attend the research, you may quit at any time. If you decide to continue, you will be asked to complete a series of formative assessment/motivation/achievement related activities during the whole field work. For example:

1. Attend the English lessons as usual;
2. Complete 2 questionnaires in **Week 1 and Week 10** in the main study;
3. Complete 2 tests in **Week 5 and Week 10** in the main study;
4. Complete a quiz on each Friday for 10 weeks;
5. Attend individual interview or group interview if being asked.

The interviews will be recorded, but all data will only be used for PhD thesis, or the presentation at academic conferences. **Your information will not be recognised in any publications.**

By joining this research, you may develop your English, and get to know the use of formative assessment and its impact on learners' motivation and achievement. However, you may feel a little uncomfortable about the questionnaire, test, and quiz.
The research study will not stop unless force majeure happens. In that case, once overcome the difficulty, the study will start over from the beginning. All the data from previous study will be deleted and destroyed. The study will continue with totally different participants who will be selected with the same method.

This project has been ethically approved via the School of Education's ethics review procedure. If something goes wrong, you can contact me whenever you want. If not, you can report your worry to the head teacher of your school, or my supervisor: Tim Herrick. If your complaint has not been handled to your satisfaction, you can contact the University's 'Registrar and Secretary'.

If you are interested in my research results, please contact me to ask for a copy of the results. If you have any other questions, please call: 18344762989; or email me.

Thank you very much for your help!

Research student: Jiawei Zhang: Jiawei.zhang@sheffield.ac.uk
Research supervisor: Tim Herrick: t.herrick@sheffield.ac.uk
Registrar and Secretary: registrar@sheffield.ac.uk

Appendix 9 Information sheet for teachers (Main study original-English)

The Impact of Formative Assessment on Learners' Motivation and Achievement in Language Learning
- A Study with English Learners in Primary Schools in China

You are being invited to take part in a research project. This research is being undertaken for a Doctor of Philosophy degree at the University of Sheffield in the UK. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask me if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

Project Overview

The aim of this study is to investigate the impact of formative assessment on learners' motivation and achievement in the language learning. This project starts from March to May in 2015, including a pilot study (2 weeks) and a main study (10 weeks). This information sheet discusses the issues in the **Main Study.**

Two public schools in Jiangsu Province with similar status will be selected. After looking across the abilities of classes, you and your class are reached. There are four classes will be invited in total, two from each school. If you are interested in my study, please discuss this study with your students who are learning English with you. If both you and your class are interested in and willing to take part in this study, you and your class will be chosen as participants for my research.

You will take part in an interview, designing two examinations and 10 quizzes. Most importantly, please use as much formative assessment as you can. If you choose to take part in this study, I will be asking you to complete a series of formative and reflective tasks over the next 10 weeks. There will include using formative assessment in your English classes more often; having an interview on formative assessment, students' motivation and English achievement; designing and marking two examinations in Week 5 and Week 10; and prepare a quiz for each week in the main study. Meanwhile, the learners are expected to have 2 exams and 10 quizzes accordingly. Some of them will take part in the interviews too.

As an English teacher, you are being asked to take part in this research.

Your Participation

Your participation in this research is entirely voluntary at all stages of the process. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a consent form. If, at any stage of the process you decide to withdraw, you are free to do so. You do not have to give a reason.

Your participation in the research will involve:

(a) You will be asked to use more formative assessment in the English class. The formative assessment is defined as sharing learning goals, questioning, feedback, and peer/self-assessment, in this research, which are common teaching and assessing techniques. Please use as much as you can

but only if it will not disturb your usual teaching routine and will not get in the way of students' learning.

(b) You will also be invited to participate in an interview. This will last approximately one hour and will be recorded. All interviews will take place in your school at a time that is convenient to you. The interviews will comprise a selection of general and focused questions that will enable you to talk about your perception about formative assessment and its impact on students' motivation and achievement. There are no right and wrong answers. You do not need to prepare for the interview in any way.

(c) You will be asked to design and mark an exam paper in Week 5 and Week 10 respectively and prepare a quiz for each week. The exams need to be marked by you, but the quizzes will be marked by students themselves.

I will make every effort to ensure that the research process is a positive experience for you. I commit to ensuring that your well-being is not compromised at any stage of the research. Should you wish to withdraw from the project, you will be free to do so. Likewise, if you feel at any stage of the project that you wish to raise an issue, you may contact my doctoral supervisor (Tim Herrick). If your complaint has not been handled to your satisfaction, you can contact the University's 'Registrar and Secretary'.

The results will show the impact of formative assessment on students' motivation and achievement. They will be shared with participating schools, so there is an opportunity for developments to occur within the teacher's immediate context. Also, it is hoped that this work will awake your awareness of the use of formative assessment. According to the results, you and other English teachers may use these techniques more wisely. Hopefully, by doing so, this research will help develop language learning in China.

USE OF DATA AND CONFIDENTIALITY

Only the interview will be recorded with audio recorder. Data from this research project will be used for the purposes of preparing a doctoral thesis for the University of Sheffield. No other use will be made of them without your written permission, and no one outside the project will be allowed access to the original recordings. The findings may also be used in preparation of journal articles or conference papers relating to the doctoral thesis in the future. The proposed date for completion of the doctoral thesis is April 2017. When the results are likely to be published, you can email me to obtain a copy of the published results.

The research study will not stop unless force majeure happens. In that case, once overcome the difficulty, the study will start over from the beginning. All the data from previous study will be deleted and destroyed. The study will continue with totally different participants who will be selected with the same method.

All information collected over the course of the research will be kept strictly confidential. Data will be stored on an external hard drive which will be encrypted for protection. Your real name will not be used in the thesis or in any subsequent report or publication.

ETHICS

This project has been ethically approved via the School of Education's ethics review procedure. The University's Research Ethics Committee monitors the application and delivery of the University's Ethics Review Procedure across the University.

I would like to thank you for taking the time to consider participating in this research project. If you wish to discuss any aspect of the research in advance of providing consent, please contact me via details provided below:

Research student: Jiawei Zhang: Jiawei.zhang@sheffield.ac.uk
Research supervisor: Tim Herrick: t.herrick@sheffield.ac.uk
Registrar and Secretary: registrar@sheffield.ac.uk

Appendix 10 Consent form for students (Main study original-Chinese)

同意书

注：此研究项目仅作英国谢菲尔德大学教育学博士学位论文撰写之用途。

项目名称：形成性评价在中国小学生英语学习过程中对学生动机/成就的影响

研究员：张家维

参与者：

请和你的家长讨论信息函的内容，如果你同意以下各项，请你的家长在方格中打勾

☐ 我确认我读完关于以上项目的信息书（时间为 2015-03-16），并且我有机会提出相关问题。

☐ 我明白我可以随时退出此次项目。当我想退出时，我可以打电话到 18344762989 进行确认。

☐ 我明白在研究过程中，一切会匿名进行。研究者可以接触到我匿名后提供的信息。

☐ 我愿意参加此次研究。

☐ 我同意在被采访时录音。

☐ 我同意我所提出的信息可以被匿名引用。

_____	_____	_____
学生姓名（拼音）	日期	家长签名
_____	_____	_____
研究者	日期	签名

请在参与者在场的情况下签字。

Appendix 10.1 Consent form for students (Main study translated-English)

Participant Consent Form

Note: this research is being undertaken for a Doctor of Philosophy degree at the University of Sheffield in the UK.

Title of Project: The Impact of Formative Assessment on Learners' Motivation and Achievement in Language Learning - A Study with English Learners in Primary Schools in China

Name of Researcher: Jiawei Zhang

Participant Identification Number for this project:

Please have a discussion with your parents or guardians about the information sheet, please let your parents or guardians tick the box and sign the form if you agree with the contents

- ☐ I confirm that I have read and understand the information sheet dated 16/03/2015 for the above project and have had the opportunity to ask questions.
- ☐ I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason. When I am about to drop out the project, I can call 18344762989.
- ☐ I understand that my responses will be anonymised before analysis. I give permission for members of the research team to have access to my anonymised responses.
- ☐ I agree to take part in the above study.
- ☐ I agree to the interview being audio recorded.
- ☐ I agree to the use of anonymised quotes in publications.

_____	_____	_____
Name of Participant (Pinyin)	Date	Signature of Participant's Guardian
_____	_____	_____
Researcher	Date	Signature

To be signed and dated in presence of the participant.

Appendix 11 Consent form for teachers (Main study original-English)

Participant Consent Form

Note: this research is being undertaken for a Doctor of Philosophy degree at the University of Sheffield in the UK.

Title of Project: The Impact of Formative Assessment on Learners' Motivation and Achievement in Language Learning -A Study with English Learners in Primary Schools in China

Name of Researcher: Jiawei ZHANG

Participant Identification Number for this project:

Please tick the box if you agree

- ☐ I confirm that I have read and understand the information sheet dated 16/03/2015 for the above project and have had the opportunity to ask questions.
- ☐ I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason. When I am about to drop out the project, I can call 18344762989.
- ☐ I understand that my responses will be anonymised before analysis.
- ☐ I give permission for members of the research team to have access to my anonymised responses.
- ☐ I agree to take part in the above study.
- ☐ I agree to the interview being audio recorded.
- ☐ I agree to the use of anonymised quotes in publications.

_____	_____	_____
Name of Participant	Date	Signature
_____	_____	_____
Researcher	Date	Signature

To be signed and dated in presence of the participant.

Appendix 12 Interview transcripts sample (Main study original-Chinese)

记：感谢参加本次采访，所有的影像资料都不会被传播出去，这次录音只是为了日后分析文字稿是精确地分析。请问如果继续录音的话，可以吗？

郜：可以的。

记：我们这次访谈的主题就是教师对形成性评价的一些看法，开始的时候会问一些比较泛的问题，后来会根据不同的形成性评价的策略问一些细的问题。首先请问老师您对形成性评价有一个什么样的整体认知？

郜：我觉得评价是对某一个时间段行为的一种肯定或者否定，或者是一种指正。我觉得对小孩子来说，在一个特定的过程时间内，形成型的评价对他们是很重要的。因为这些评价其实就是老师观点的一种认可，一种传输，那么对孩子来说，他知道老师想要什么，那么他才可以从老师想要的这个方向去努力。因为小学生还是他我自律，他还没有达到那种用自我来约束自己，他只能通过别人的肯定/否定，以及同伴的肯定/否定，来修正自己的行为，让自己变得更加社会化，这是我的个人感觉。

记：在我长期的课堂观察中发现，老师您用的形成性评价非常多，在你心目中哪一些策略或者教学手段是可以算作形成性评价的呢？

郜：我觉得形成性评价他的方式有很多中，就语言来说，他可能是有声语言也有可能是无声语言，比如说我可能会通过一个 great, wonderful，同时用我的目光，来表示共鸣我的评价。那么其实他还有无声的语言，你可能会发现，我会用哪怕一秒钟的沉默来传递我的观点。那么第二个来说，有声语言无声语言，还有身体语言，body language，我觉得 body language 也是一种评价。这些都是适合小孩子的。比如说老师摸一下头，摸一下小手，拥抱一下，这些方式都可能透露老师对孩子 ，无论是赞美的是批评的，是肯定的是否定的，其实这些 body language 只是告诉小孩子老师可能只是不喜欢你的某一个行为，但不是不喜欢你这个人。我觉得评价本身只是建立在爱与理解之上，所产生的一种对行为的评价方式。

Appendix 12.1 Interview transcripts sample (Main study translated-English)[3]

Me: thanks to participate in this interview, all the information will not be spread out, the recording is only for the future accurate text analysis. Would you like to continue?

RT: Yes.

Me: The theme of our interview is the teacher's view of formative assessment, in the beginning, I will ask some of the more general questions, and later will ask some fine questions on the basis of different FA strategies. Firstly, what do you think of FA in general?

RT: I think assessment is a confirmation or criticise on behaviour during a certain period of time, or a correction. I think for children, in a specific period of time, FA is very important to them. Because these assessments are actually a recognition, or a transmission of teacher's viewpoint, then to the children, they know what the teacher wants, then they can work hard in the direction according to what the teacher wants. Because the primary school students are mainly disciplined by others, they have not reached the kind of self-discipline with their own constraints, they can only amend their behaviour via affirmation or denial by others, and peer affirmation or denial, to make themselves more socialised. This is my personal perspective.

Me: In my classroom observation for a long time, I found that you use FA a lot, in your mind which strategies or teaching methods can be counted as FA?

RT: I think there are many strategies in FA, in terms of the language[4], it may be a voice language, may also be a silent language, for example, I may show my assessment via a *"great", "wonderful"*[5], with my *eye-contact* at the same time. There are some other silent languages, you may find that I will use even a second of silence to pass my viewpoint. Secondly, the voice language, silent language, as well as body language, body language, I think the body language is also an assessment. These are all suitable for children. For example, the *teacher touches children's head*, *holds a small hand*, *hugs*, these methods may **reveal the teacher's affection**[6] to the children, whether it is *praise* or *criticise*, is positive or negative, in fact, these body languages just tell the children that **the teacher may just not like one of your behaviours, but not dislike you as a person**. I believe that the assessment itself is built on the basis of **love and understanding**, then produced naturally to **assess the behaviours**.

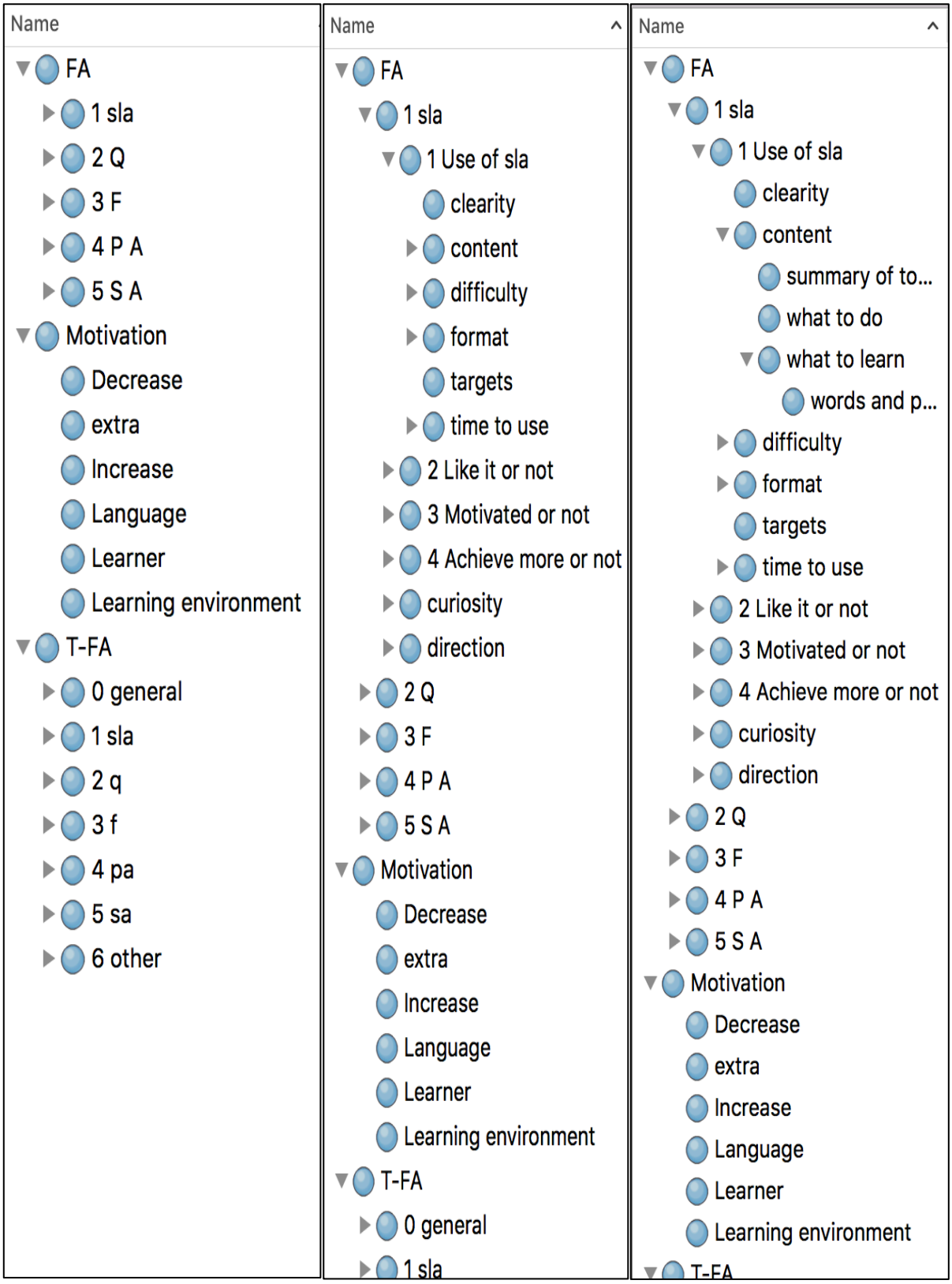
[3] The codes in this sample are examples that captured from NVivo, which are in a different form

[4] Underline: the categorises of FA strategies

[5] Italic: the concrete examples of FA strategies

[6] Bold: the reasons why teacher using FA

Appendix 12.2 The structure of codes



Appendix 12.3 Codes compared by number of coding references

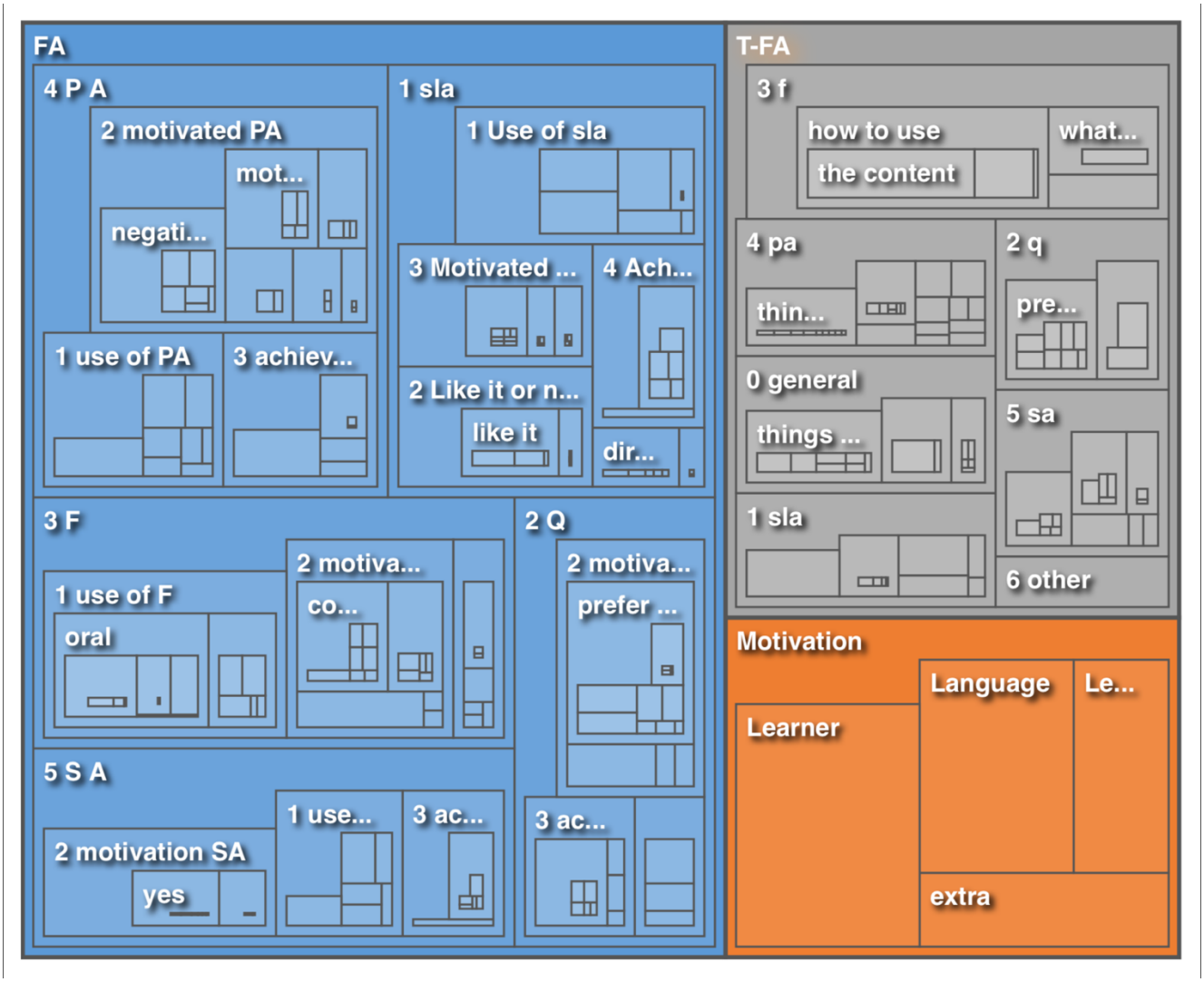


Table 1 Paired Samples Test of Motivation Questionnaire

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	M1	-.041	.960	.069	-.176	.094	-.595	195	.552
Pair 2	M2	-.031	.919	.066	-.161	.099	-.468	194	.641
Pair 3	M3	.015	1.020	.073	-.129	.160	.211	193	.833
Pair 4	M4	-.093	1.066	.077	-.245	.058	-1.215	192	.226
Pair 5	M5	.066	1.393	.099	-.130	.262	.667	195	.506
Pair 6	M6	-.052	1.106	.080	-.209	.105	-.653	191	.515
Pair 7	M7	.130	1.690	.122	-.110	.371	1.067	191	.287
Pair 8	M8	.005	1.274	.091	-.175	.185	.056	194	.955
Pair 9	M9	-.102	.764	.055	-.210	.006	-1.869	195	.063
Pair 10	M10	-.117	.912	.065	-.246	.011	-1.801	195	.073
Pair 11	M11	-.103	1.074	.077	-.254	.049	-1.333	194	.184
Pair 12	M12	.000	.987	.071	-.140	.140	.000	193	1.000
Pair 13	M13	.113	1.417	.102	-.087	.314	1.115	193	.266
Pair 14	M14	-.072	1.089	.078	-.226	.082	-.923	193	.357
Pair 15	M15	.000	1.167	.084	-.166	.166	.000	191	1.000
Pair 16	M16	.264	1.805	.130	.008	.521	2.034	192	.043
Pair 17	M17	.031	1.626	.117	-.200	.262	.266	192	.791
Pair 18	M18	.021	1.593	.114	-.205	.246	.180	193	.857
Pair 19	M19	-.129	1.619	.116	-.358	.100	-1.109	193	.269
Pair 20	M21	.113	1.474	.106	-.095	.321	1.069	194	.286

Table 2 Component Matrix[a]

	Component			
	1	2	3	4
M11: I have high faith in my English	.857			
M12: I have strong faith in my English	.836			
M9: I think I have the ability to learn English well	.805			
M3: English is useful to me	.761			
M4: I'm willing to spend time on learning English	.748			
M2: English is interesting to me	.736			
M8: I learn English because I want to master this language	.724			
M10: I think I work hard in learning English	.720			
M14: Learning English successfully is what I want	.720			
M5: Learning aims are easy to me	.663			.436
M1: English is important to me	.657			
M6: Learning aims are clear to me	.645			.447
M21: I learn English because I like the western culture	.611			
M15: I like the feeling of success	.555			-.460
M18: My families have great impact on my English		.840		
M17: My teacher has great impact on my English	.401	.780		

M19: My friends have great impact on my English	.436	.756	
M16: I am afraid of the feeling of failure		.496	.419
M7: I learn English because I want to be better than others			.492
M13: Learning English successfully is what others want me to achieve	.465		.486

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

Table 3 Pattern Matrix[a]

	Component			
	1	2	3	4
M4: I'm willing to spend time on learning English	**.836**			
M5: Learning aims are easy to me	**.790**			
M6: Learning aims are clear to me	**.783**			
M2: English is interesting to me	**.777**			
M3: English is useful to me	**.588**			
M9: I think I have the ability to learn English well	**.474**	**.466**		
M21: I learn English because I like the western culture				
M15: I like the feeling of success		*.893*		
M1: English is important to me		*.718*		
M8: I learn English because I want to master this language		*.679*		
M14: Learning English successfully is what I want		*.675*		
M12: I have strong faith in my English		*.556*		
M11: I have high faith in my English	*.463*	*.483*		
M10: I think I work hard in learning English		*.470*		
M18: My families have great impact on my English			.930	
M17: My teacher has great impact on my English			.918	
M19: My friends have great impact on my English			.909	
M7: I learn English because I want to be better than others				.697
M13: Learning English successfully is what others want me to achieve				.630
M16: I am afraid of the feeling of failure				.622

Extraction Method: Principal Component Analysis.
Rotation Method: Promax with Kaiser Normalization.
a. Rotation converged in 17 iterations.

Table 4 Correlation between FA and motivation in different achievement groups

			Pre-test				Post-test			
			Language	Learner	Learning	Motivation	Language	Learner	Learning	Motivation
LA	SLA	r	**.721****	**.656****	.317	**.691****	**.520***	**.556***	.355*	**.583****
		p	.000	.000	.082	.000	.002	.001	.039	.000
	Q-O	r	**.583****	**.603****	.204	**.602****	**.565****	.462*	.322	.496*
		p	.001	.000	.270	.000	.000	.006	.063	.003
	Q-C	r	**.516***	**.514***	.293	**.539***	**.519***	**.514***	.260	**.570***
		p	.003	.003	.110	.002	.002	.002	.138	.000
	F-CF	r	.411*	.467*	.181	.448*	.471*	.368*	.271	.383*
		p	.022	.008	.331	.011	.005	.032	.121	.026
	F-S	r	**.700****	**.674****	.220	**.668****	**.570****	**.515***	.237	**.522***
		p	.000	.000	.235	.000	.000	.002	.177	.002
	F-CR	r	.442*	**.514***	.323	**.526***	.323	.316	-.018	.311
		p	.013	.003	.076	.002	.062	.069	.922	.073
	PA-P	r	**.776****	**.703****	.308	**.699****	**.633****	**.526****	**.596****	**.628****

		p	.000	.000	.092	.000	.000	.001	.000	.000
	PA-N	r	-.382*	-.432*	-.345	-.416*	.437*	.452*	.226	.483*
		p	.034	.015	.058	.020	.010	.007	.198	.004
	sa	r	.435*	.366*	.219	.369*	**.618****	**.570****	**.536****	**.637****
		p	.014	.043	.237	.041	.000	.000	.001	.000
MA	SLA	r	.305**	.299**	-.013	.278*	.267*	.235*	.031	.279*
		p	.001	.001	.885	.002	.003	.009	.731	.002
	Q-O	r	.301**	.341**	-.004	.246*	.352**	.272*	.007	.323**
		p	.001	.000	.965	.007	.000	.002	.935	.000
	Q-C	r	.168	.204*	-.055	.122	.266*	.210*	.020	.239*
		p	.066	.025	.549	.183	.003	.019	.829	.008
	F-CF	r	.194*	.168	.002	.088	.311**	.252*	.190*	.280*
		p	.033	.066	.979	.340	.000	.005	.035	.002
	F-S	r	.279*	.302**	-.068	.204*	.306**	.266*	.165	.342**
		p	.002	.001	.459	.025	.001	.003	.067	.000
	F-CR	r	.333**	.379**	.037	.307**	.069	.070	.235*	.232*
		p	.000	.000	.684	.001	.445	.442	.009	.010
	PA-P	r	.299**	.291**	.028	.221*	.355**	.340**	.077	.356**
		p	.001	.001	.763	.015	.000	.000	.395	.000
	PA-N	r	-.018	.015	.228*	.129	.007	.050	.174	.179*
		p	.849	.873	.012	.158	.937	.578	.054	.047
	sa	r	.386**	.342**	.093	.293**	.388**	.455**	.150	.448**
		p	.000	.000	.312	.001	.000	.000	.096	.000
HA	SLA	r	.434*	.361	.113	.409*	.266	.069	.417*	.355
		p	.021	.059	.567	.031	.172	.729	.027	.063
	Q-O	r	.470*	.321	.119	.266	.175	.003	.195	.154
		p	.012	.096	.545	.171	.374	.989	.321	.433
	Q-C	r	.386*	.450*	.147	.369	**.522***	.223	.474*	.442*
		p	.043	.016	.455	.053	.004	.253	.011	.019
	F-CF	r	.118	.198	.141	.180	.479*	.105	**.512***	.432*
		p	.548	.312	.473	.360	.010	.596	.005	.022
	F-S	r	.419*	.464*	.138	.325	**.518***	.203	.352	.408*
		p	.026	.013	.484	.091	.005	.300	.066	.031
	F-CR	r	.462*	.371	.248	.422*	.257	.367	.273	.407*
		p	.013	.052	.204	.025	.187	.055	.159	.032
	PA-P	r	**.750****	**.618****	.269	**.554***	**.739****	**.590****	**.632****	**.689****
		p	.000	.000	.166	.002	.000	.001	.000	.000
	PA-N	r	-.391*	-.464*	.057	-.188	.125	.095	.492*	.320
		p	.040	.013	.772	.337	.527	.629	.008	.097
	sa	r	**.660****	**.515****	.442*	**.701****	**.700****	**.661****	**.673****	**.808****
		p	.000	.005	.019	.000	.000	.000	.000	.000

** Correlation is significant at the .001 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

r= Pearson Correlation

p= Sig. (2-tailed)

Table 5 Correlation between FA and motivation in different FA environments

			Pre-test				Post-test			
			Language	Learner	Learning	Motivation	Language	Learner	Learning	Motivation
FAR	SLA	r	.455**	.393**	-.008	.312*	.413**	.354**	.247*	.452**
		p	.000	.000	.938	.003	.000	.000	.017	.000
	Q-O	r	.343**	.332**	.108	.277*	**.502****	.414**	.123	.414**
		p	.001	.001	.314	.009	.000	.000	.241	.000
	Q-C	r	.412**	.435**	.051	.338**	.487**	.393**	.088	.406**

		p	.000	.000	.634	.001	.000	.000	.402	.000
	F-CF	r	.337**	.324*	.094	.256*	**.594****	.494**	.209*	.497**
		p	.001	.002	.383	.015	.000	.000	.045	.000
	F-S	r	.357**	.414**	.085	.286*	**.515****	.444**	.267*	**.502****
		p	.001	.000	.429	.006	.000	.000	.010	.000
	F-CR	r	.398**	.424**	.141	.393**	.083	.112	.147	.225*
		p	.000	.000	.188	.000	.431	.287	.160	.030
	PA-P	r	**.538****	.484**	.206	.448**	**.501****	**.512****	.238*	**.516****
		p	.000	.000	.052	.000	.000	.000	.022	.000
	PA-N	r	-.239*	-.221*	.083	-.118	.059	.129	.278*	.273*
		p	.024	.038	.440	.269	.577	.218	.007	.008
	sa	r	.399**	.364**	.164	.322*	.363**	.467**	.364**	**.543****
		p	.000	.000	.125	.002	.000	.000	.000	.000
FA	SLA	r	**.537****	**.519****	-.003	**.525****	.270*	.248*	.071	.345**
L		p	.000	.000	.979	.000	.008	.016	.499	.001
	Q-O	r	**.509****	**.546****	-.099	.443**	.434**	.315*	.010	.369**
		p	.000	.000	.349	.000	.000	.002	.926	.000
	Q-C	r	.364**	.380**	-.081	.311*	.233*	.191	.148	.283*
		p	.000	.000	.442	.003	.024	.065	.154	.006
	F-CF	r	.343**	.359**	-.024	.267*	.358**	.239*	.200	.272*
		p	.001	.000	.821	.010	.000	.020	.053	.008
	F-S	r	**.550****	**.535****	-.063	.477**	.443**	.354**	.119	.428**
		p	.000	.000	.550	.000	.000	.000	.255	.000
	F-CR	r	.343**	.384**	.073	.370**	.252*	.238*	.298*	.382**
		p	.001	.000	.490	.000	.014	.021	.004	.000
	PA-P	r	.407**	.389**	-.062	.312*	.479*	.378*	.176	.463*
		p	.000	.000	.560	.002	.000	.000	.089	.000
	PA-N	r	-.020	-.035	.151	.057	.159	.150	.175	.282*
		p	.851	.743	.150	.592	.126	.150	.092	.006
	sa	r	.462**	.385**	.059	.399**	**.636****	**.572****	.112	**.555****
		p	.000	.000	.578	.000	.000	.000	.282	.000

** Correlation is significant at the .001 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

r= Pearson Correlation

p= Sig. (2-tailed)

Table 6 Correlation between FA and motivation in different genders

			Pre-test				Post-test			
			Language	Learner	Learning	Motivation	Language	Learner	Learning	Motivation
B	SLA	r	**.565****	**.532****	.063	.497**	.187	.212*	.179	.290**
		p	.000	.000	.555	.000	.068	.038	.081	.004
	Q-O	r	**.501****	**.542****	.098	.476**	.460**	.335*	.017	.334*
		p	.000	.000	.355	.000	.000	.001	.871	.001
	Q-C	r	.475**	**.534****	.040	.420**	.406**	.315*	.096	.380**
		p	.000	.000	.708	.000	.000	.002	.353	.000
	F-CF	r	.395**	.411**	.051	.328**	**.501****	.383**	.272*	.430**
		p	.000	.000	.634	.001	.000	.000	.007	.000
	F-S	r	**.508****	**.516****	-.023	.434**	.418**	.302*	.180	.383**
		p	.000	.000	.829	.000	.000	.003	.079	.000
	F-CR	r	.302*	.350**	.153	.389**	.136	.126	.150	.244*
		p	.004	.001	.148	.000	.186	.221	.143	.017
	PA-P	r	.495**	.448**	-.039	.375**	.490**	.363**	.153	.418**
		p	.000	.000	.716	.000	.000	.000	.136	.000
	PA-N	r	-.154	-.157	.024	-.083	.126	.166	.295*	.389**

		p	.145	.138	.823	.435	.222	.106	.004	.000
	sa	r	.357**	.288*	.082	.283*	**.558****	**.542****	.306*	**.604****
		p	.001	.006	.438	.007	.000	.000	.002	.000
G	SLA	r	.416**	.400**	-.014	.381**	.420**	.328**	.087	.423**
		p	.000	.000	.898	.000	.000	.001	.413	.000
	Q-O	r	.343**	.326*	-.068	.217*	.410**	.352**	.077	.403**
		p	.001	.002	.523	.040	.000	.001	.467	.000
	Q-C	r	.260*	.228*	-.037	.195	.301*	.271*	.119	.308*
		p	.013	.031	.729	.066	.004	.009	.260	.003
	F-CF	r	.214*	.207*	.003	.122	.351**	.271*	.082	.246*
		p	.043	.050	.978	.253	.001	.009	.439	.019
	F-S	r	.399**	.410**	.046	.318*	.455**	.424**	.157	.448**
		p	.000	.000	.666	.002	.000	.000	.137	.000
	F-CR	r	.472**	.486**	.090	.386**	.164	.196	.259*	.302*
		p	.000	.000	.401	.000	.120	.063	.013	.004
	PA-P	r	.444**	.425**	.233*	.398**	.481**	**.511****	.268*	**.547****
		p	.000	.000	.027	.000	.000	.000	.010	.000
	PA-N	r	-.016	-.023	.238*	.100	.086	.100	.152	.124
		p	.880	.826	.024	.348	.415	.343	.151	.241
	sa	r	**.576****	**.532****	.207*	**.513****	.467**	**.522****	.192	**.514****
		p	.000	.000	.050	.000	.000	.000	.068	.000

** Correlation is significant at the .001 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

r= Pearson Correlation

p= Sig. (2-tailed)

Table 7 Correlation between FA and motivation in different genders and achievement groups

				Pre-test				Post-test			
				Language	Learner	Learning	Motivation	Language	Learner	Learning	Motivation
B	L	SLA	r	**.570***	.482*	.290	**.566***	.402	.499*	.319	.460*
			p	.011	.037	.228	.011	.071	.021	.158	.036
		Q-O	r	**.502***	**.526***	.278	**.555***	**.548***	.408	.286	.431
			p	.028	.021	.248	.014	.010	.067	.209	.051
		Q-C	r	.360	.382	.448	.473*	.444*	.391	.286	.479*
			p	.130	.107	.054	.041	.044	.080	.209	.028
		F-CF	r	.374	.396	.185	.420	**.560***	.444*	.277	.443*
			p	.114	.093	.449	.074	.008	.044	.225	.044
		F-S	r	**.611***	**.533***	.265	**.599***	.483*	.420	.309	.467*
			p	.005	.019	.272	.007	.026	.058	.173	.033
		F-CR	r	.358	**.476***	.323	.454	.215	.279	.063	.236
			p	.132	.039	.177	.051	.349	.220	.788	.302
		PA-P	r	**.753****	**.612***	.355	**.661***	**.680****	**.550***	**.619***	**.649****
			p	.000	.005	.136	.002	.001	.010	.003	.001
		PA-N	r	-.446	-.468*	**-.531***	-.475*	.393	.486*	.236	**.502***
			p	.056	.043	.019	.040	.078	.026	.303	.020
		sa	r	.311	.261	.199	.265	**.752****	**.705****	**.528***	**.754****
			p	.196	.280	.413	.272	.000	.000	.014	.000
	M	SLA	r	.311*	.268*	.053	.290*	.189	.218	-.016	.201
			p	.016	.038	.686	.025	.137	.085	.904	.113
		Q-O	r	.324*	.414**	.058	.316*	.366*	.255*	-.024	.296*
			p	.012	.001	.661	.014	.003	.043	.849	.018
		Q-C	r	.341*	.461**	-.055	.260*	.342*	.276*	-.005	.299*
			p	.008	.000	.678	.045	.006	.028	.969	.017

		F-CF	r	.239	.248	.044	.132	.405**	.362*	.290*	.394*
			p	.066	.056	.737	.316	.001	.004	.021	.001
		F-S	r	.292*	.341*	-.115	.217	.300*	.252*	.109	.308*
			p	.024	.008	.381	.096	.017	.047	.393	.014
		F-CR	r	.179	.238	.066	.287*	.056	.001	.173	.206
			p	.172	.067	.619	.026	.662	.994	.176	.105
		PA-P	r	.250	.243	-.210	.091	.278*	.224	.040	.257*
			p	.054	.061	.108	.490	.027	.077	.754	.042
		PA-N	r	-.035	.014	.199	.133	.042	.086	.220	.315*
			p	.790	.913	.127	.309	.741	.502	.084	.012
		sa	r	.224	.167	-.064	.092	.405**	.475**	.238	.493**
			p	.085	.202	.627	.483	.001	.000	.060	.000
	H	SLA	r	.166	.232	-.079	.076	.091	-.249	**.824***	.367
			p	.646	.519	.828	.835	.802	.488	.003	.297
		Q-O	r	**.814***	.592	.301	.500	.279	-.111	.421	.196
			p	.004	.071	.398	.141	.436	.760	.225	.588
		Q-C	r	.042	-.046	.016	-.029	.549	.002	.538	.389
			p	.909	.900	.966	.936	.101	.996	.109	.267
		F-CF	r	-.259	-.263	-.111	-.224	.630	.116	.617	.561
			p	.470	.463	.761	.533	.051	.750	.057	.091
		F-S	r	.042	-.046	.016	-.029	**.648***	.123	.606	.535
			p	.909	.900	.966	.936	.043	.736	.063	.111
		F-CR	r	.251	-.054	.461	.353	.123	.307	.240	.359
			p	.485	.882	.180	.318	.735	.388	.504	.308
		PA-P	r	**.800***	.464	.567	**.687***	**.719***	.229	.431	.518
			p	.005	.176	.088	.028	.019	.524	.214	.125
		PA-N	r	-.323	-.523	-.221	-.318	.340	.232	**.683***	**.636***
			p	.363	.120	.540	.370	.337	.519	.029	.048
		sa	r	**.672***	.243	**.777***	**.765***	.610	.345	**.711***	**.756***
			p	.033	.498	.008	.010	.061	.329	.021	.011
G	L	sla	r	**.967****	**.900****	.472	**.928****	**.672***	**.630***	.393	**.745***
			p	.000	.000	.142	.000	.017	.028	.207	.005
		Q-O	r	**.739***	**.736***	.020	**.664***	**.713***	**.671***	.528	**.744***
			p	.009	.010	.954	.026	.009	.017	.078	.005
		Q-C	r	**.713***	**.716***	.015	.585	**.699***	**.787***	.551	**.843****
			p	.014	.013	.964	.059	.012	.002	.063	.001
		F-CF	r	.469	.548	.222	.482	.269	.100	.277	.155
			p	.146	.081	.512	.133	.397	.758	.384	.630
		F-S	r	**.779***	**.838****	.149	**.727***	**.780***	**.750***	.058	**.677***
			p	.005	.001	.661	.011	.003	.005	.858	.016
		F-CR	r	.537	.539	.321	**.628***	.384	.271	-.084	.335
			p	.089	.087	.335	.039	.218	.394	.794	.286
		PA-P	r	**.790***	**.769***	.274	**.724***	**.702***	**.624***	.569	**.738***
			p	.004	.006	.414	.012	.011	.030	.053	.006
		PA-N	r	.101	.144	.149	.083	.495	.335	.191	.396
			p	.768	.673	.662	.809	.102	.288	.551	.202
		sa	r	**.685***	**.618***	.255	.585	.387	.298	**.635***	.411
			p	.020	.043	.450	.059	.214	.346	.027	.185
	M	SLA	r	.283*	.328*	-.067	.267*	.347*	.248	.083	.357*
			p	.028	.010	.611	.039	.007	.056	.528	.005
		Q-O	r	.272*	.278*	-.060	.178	.359*	.304*	.046	.369*
			p	.036	.031	.646	.174	.005	.018	.728	.004
		Q-C	r	.048	.009	-.055	.025	.199	.149	.044	.183
			p	.715	.946	.679	.847	.128	.257	.740	.161
		F-CF	r	.144	.101	-.030	.046	.224	.148	.090	.167
			p	.274	.445	.818	.728	.085	.260	.492	.201

	F-S	r	.256*	.270*	-.012	.186	.325*	.295*	.216	.390*
		p	.048	.037	.927	.154	.011	.022	.098	.002
	F-CR	r	.481**	.487**	.023	.333*	.091	.145	.301*	.270*
		p	.000	.000	.859	.009	.489	.268	.019	.037
	PA-P	r	.345*	.336*	.269*	.342*	.435**	.454**	.115	.456**
		p	.007	.009	.037	.007	.001	.000	.381	.000
	PA-N	r	.006	.015	.262*	.129	-.023	.016	.127	.038
		p	.966	.911	.043	.328	.864	.904	.332	.772
	sa	r	**.553****	.496**	.259*	.492**	.386*	.448**	.065	.416**
		p	.000	.000	.045	.000	.002	.000	.623	.001
H	SLA	r	.472	.362	.021	.392	.473	.257	.060	.329
		p	.055	.154	.935	.120	.055	.319	.818	.197
	Q-O	r	.261	.161	-.018	.064	.038	.038	.015	.078
		p	.312	.537	.945	.807	.884	.885	.954	.765
	Q-C	r	**.594***	**.722****	.242	**.686***	.491*	.381	.331	.418
		p	.012	.001	.350	.002	.045	.132	.194	.095
	F-CF	r	.393	**.517***	.338	**.539***	.089	.138	.377	.251
		p	.119	.033	.185	.026	.733	.597	.136	.332
	F-S	r	**.532***	**.626***	.107	.412	.315	.267	.002	.228
		p	.028	.007	.683	.101	.217	.301	.994	.379
	F-CR	r	**.584***	**.602***	.075	.474	.444	.436	.357	.499*
		p	.014	.011	.775	.055	.074	.080	.159	.041
	PA-P	r	**.754****	**.674***	.034	.476	**.826****	**.785****	**.838****	**.846****
		p	.000	.003	.898	.053	.000	.000	.000	.000
	PA-N	r	-.422	-.418	.314	-.027	-.073	.021	.343	.093
		p	.092	.095	.220	.918	.779	.935	.177	.723
	sa	r	**.630***	**.649***	.211	**.635***	**.834****	**.899****	**.620***	**.851****
		p	.007	.005	.415	.006	.000	.000	.008	.000

** Correlation is significant at the .001 level (2-tailed).
* Correlation is significant at the .05 level (2-tailed).
r= Pearson Correlation
p= Sig. (2-tailed)

Table 8 Correlation between use of FA and achievement in different achievement groups

		LA	MA	HA
Use of Sharing Learning Aims	Pearson Correlation	.344*	-.120	.154
	Sig. (2-tailed)	.043	.176	.425
Use of Open-ended Questions	Pearson Correlation	.266	.068	.263
	Sig. (2-tailed)	.122	.447	.169
Use of Close-ended Questions	Pearson Correlation	.298	.031	.321
	Sig. (2-tailed)	.083	.726	.089
Use of Confirmative Feedback	Pearson Correlation	.033	.142	.296
	Sig. (2-tailed)	.849	.107	.119
Use of Suggestive Feedback	Pearson Correlation	.049	-.017	.215
	Sig. (2-tailed)	.781	.845	.264
Use of Corrective Feedback	Pearson Correlation	.415*	-.017	-.058
	Sig. (2-tailed)	.013	.849	.766
Use of Positive Peer-assessment	Pearson Correlation	.109	.240*	.285
	Sig. (2-tailed)	.533	.006	.134

Use of Negative Peer-assessment	Pearson Correlation	-.040	.009	-.106
	Sig. (2-tailed)	.822	.915	.586
Use of Self-assessment	Pearson Correlation	.297	.057	.079
	Sig. (2-tailed)	.083	.522	.684

** Correlation is significant at the .001 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

Table 9 Correlation between use of FA and achievement in different FA environments

		FAR	FAL
Use of Sharing Learning Aims	Pearson Correlation	.177	.172
	Sig. (2-tailed)	.082	.093
Use of Open-ended Questions	Pearson Correlation	.279*	.312*
	Sig. (2-tailed)	.005	.002
Use of Close-ended Questions	Pearson Correlation	.359**	.243*
	Sig. (2-tailed)	.000	.017
Use of Confirmative Feedback	Pearson Correlation	.402**	.212*
	Sig. (2-tailed)	.000	.038
Use of Suggestive Feedback	Pearson Correlation	.251*	.194
	Sig. (2-tailed)	.013	.059
Use of Corrective Feedback	Pearson Correlation	.134	.238*
	Sig. (2-tailed)	.188	.020
Use of Positive Peer-assessment	Pearson Correlation	.209*	.272*
	Sig. (2-tailed)	.039	.007
Use of Negative Peer-assessment	Pearson Correlation	.039	.044
	Sig. (2-tailed)	.703	.667
Use of Self-assessment	Pearson Correlation	.183	.307*
	Sig. (2-tailed)	.071	.002

** Correlation is significant at the .001 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

Table 10 Correlation between use of FA and achievement in different genders

		Boy	Girl
Use of Sharing Learning Aims	Pearson Correlation	.212*	.116
	Sig. (2-tailed)	.036	.263
Use of Open-ended Questions	Pearson Correlation	.308*	.271*
	Sig. (2-tailed)	.002	.008
Use of Close-ended Questions	Pearson Correlation	.381**	.134
	Sig. (2-tailed)	.000	.196
Use of Confirmative Feedback	Pearson Correlation	.181	.383**
	Sig. (2-tailed)	.073	.000
Use of Suggestive Feedback	Pearson Correlation	.200*	.195
	Sig. (2-tailed)	.047	.058

Use of Corrective Feedback	Pearson Correlation	.221*	.179
	Sig. (2-tailed)	.028	.083
Use of Positive Peer-assessment	Pearson Correlation	.260*	.217*
	Sig. (2-tailed)	.009	.035
Use of Negative Peer-assessment	Pearson Correlation	-.051	.184
	Sig. (2-tailed)	.616	.075
Use of Self-assessment	Pearson Correlation	.236*	.301*
	Sig. (2-tailed)	.019	.003

** Correlation is significant at the .001 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

Table 11 Correlation between use of FA and achievement in different genders and achievement groups

		Boy			Girl		
		LA	MA	HA	LA	MA	HA
SLA	Pearson Correlation	.104	-.062	.115	**.692***	-.195	.174
	Sig. (2-tailed)	.655	.625	.736	.009	.125	.503
Q-O	Pearson Correlation	.014	.073	.500	**.763***	.064	-.036
	Sig. (2-tailed)	.951	.561	.118	.002	.617	.892
Q-C	Pearson Correlation	.211	.197	**.604***	.478	-.107	.049
	Sig. (2-tailed)	.359	.115	.049	.098	.406	.853
F-CF	Pearson Correlation	-.189	.107	.487	**.608***	.167	.177
	Sig. (2-tailed)	.412	.395	.129	.027	.190	.498
F-S	Pearson Correlation	-.303	.025	.552	**.641***	-.067	-.174
	Sig. (2-tailed)	.182	.841	.078	.018	.603	.504
F-CR	Pearson Correlation	.257	-.125	-.262	**.594***	.114	.187
	Sig. (2-tailed)	.260	.320	.436	.032	.375	.472
PA-P	Pearson Correlation	-.118	.204	.339	.535	.267*	.150
	Sig. (2-tailed)	.612	.103	.308	.060	.034	.565
PA-N	Pearson Correlation	-.236	.145	-.306	.455	-.183	.244
	Sig. (2-tailed)	.302	.247	.360	.118	.152	.345
sa	Pearson Correlation	.138	-.063	-.156	.545	.186	.370
	Sig. (2-tailed)	.552	.620	.646	.054	.144	.144

** Correlation is significant at the .001 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

Table 12 Correlation between motivation and achievement in different achievement groups

			LA	MA	HA
Pre-test	Language	Pearson Correlation	.434*	.198*	.218
		Sig. (2-tailed)	.009	.024	.256
	Learner	Pearson Correlation	.480*	.276*	.161

		Sig. (2-tailed)	.004	.002	.403
	Learning environment	Pearson Correlation	.432*	.085	.232
		Sig. (2-tailed)	.010	.341	.226
	Motivation	Pearson Correlation	**.520****	.228*	.224
		Sig. (2-tailed)	.001	.009	.243
Post-test	Language	Pearson Correlation	.321	.207*	.360
		Sig. (2-tailed)	.060	.019	.055
	Learner	Pearson Correlation	.281	.287**	.123
		Sig. (2-tailed)	.102	.001	.526
	Learning environment	Pearson Correlation	.209	.044	.235
		Sig. (2-tailed)	.227	.620	.220
	Motivation	Pearson Correlation	.296	.215*	.133
		Sig. (2-tailed)	.084	.014	.493

** Correlation is significant at the .001 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

Table 13 Correlation between motivation and achievement in different FA environments

			FAR	FAL
Pre-test	Language	Pearson Correlation	.335**	**.536****
		Sig. (2-tailed)	.001	.000
	Learner	Pearson Correlation	.374**	**.566****
		Sig. (2-tailed)	.000	.000
	Learning environment	Pearson Correlation	-.056	.030
		Sig. (2-tailed)	.587	.770
	Motivation	Pearson Correlation	.240*	.457**
		Sig. (2-tailed)	.017	.000
Post-test	Language	Pearson Correlation	.316*	.398**
		Sig. (2-tailed)	.002	.000
	Learner	Pearson Correlation	.400**	.375**
		Sig. (2-tailed)	.000	.000
	Learning environment	Pearson Correlation	-.065	-.100
		Sig. (2-tailed)	.525	.334
	Motivation	Pearson Correlation	.208*	.236*
		Sig. (2-tailed)	.040	.020

** Correlation is significant at the .001 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

Table 14 Correlation between motivation and achievement in different genders

			boy	girl
Pre-test	Language	Pearson Correlation	.475**	.435**

	Learner	Sig. (2-tailed)	.000	.000
		Pearson Correlation	**.584****	.363**
	Learning environment	Sig. (2-tailed)	.000	.000
		Pearson Correlation	.058	-.017
	Motivation	Sig. (2-tailed)	.566	.871
		Pearson Correlation	.463**	.260*
		Sig. (2-tailed)	.000	.011
Post-test	Learner	Pearson Correlation	.334**	.435**
		Sig. (2-tailed)	.001	.000
	Language	Pearson Correlation	.355**	.432**
		Sig. (2-tailed)	.000	.000
	Learning environment	Pearson Correlation	-.105	-.011
		Sig. (2-tailed)	.299	.916
	Motivation	Pearson Correlation	.170	.326**
		Sig. (2-tailed)	.092	.001

** Correlation is significant at the .001 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

Table 15 Correlation between motivation and achievement in different achievement groups and genders

			Boy			Girl		
			LA	MA	HA	LA	MA	HA
Pre-test	Language	r	.277	.144	.384	**.700***	.238	-.065
		p	.223	.254	.244	.008	.060	.805
	Learner	r	.363	.344*	.332	**.755***	.229	-.102
		p	.105	.005	.318	.003	.071	.696
	Learning Environment	r	.452*	.168	.184	.410	-.016	.184
		p	.040	.182	.589	.164	.903	.479
	Motivation	r	.418	.269*	.229	**.712***	.168	.035
		p	.059	.030	.499	.006	.188	.893
Post-test	Language	r	.156	.213	.341	.527	.237	.479
		p	.501	.088	.304	.064	.061	.052
	Learner	r	.230	.253*	-.052	.343	.340*	.313
		p	.316	.042	.879	.251	.006	.222
	Learning Environment	r	.001	.183	.144	**.605***	-.096	.501*
		p	.996	.144	.673	.028	.455	.041
	Motivation	r	.178	.230	-.064	.457	.217	.473
		p	.439	.066	.851	.116	.087	.055

** Correlation is significant at the .001 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

r= Pearson Correlation

p= Sig. (2-tailed)