THE UNIVERSITY OF WARWICK

Library Declaration and Deposit Agreement

1. **STUDENT DETAILS**

Full name: Christopher James Cookson
University ID number: [redacted]

2. **THESIS DEPOSIT**

2.1 I understand that under my registration at the University, I am required to deposit my thesis with the University in BOTH hard copy and in digital format. The digital version should normally be saved as a single pdf file.

2.2 The hard copy will be housed in the University Library. The digital version will be deposited in the University's Institutional Repository (WRAP). Unless otherwise indicated (see 2.3 below) this will be made openly accessible on the Internet and will be supplied to the British Library to be made available online via its Electronic Theses Online Service (EThOS) service.
[At present, theses submitted for a Master's degree by Research (MA, MSc, LLM, MS or MMedSci) are not being deposited in WRAP and not being made available via EthOS. This may change in future.]

2.3 In exceptional circumstances, the Chair of the Board of Graduate Studies may grant permission for an embargo to be placed on public access to the hard copy thesis for a limited period. It is also possible to apply separately for an embargo on the digital version. (Further information is available in the *Guide to Examinations for Higher Degrees by Research*.)

2.4 *If you are depositing a thesis for a Master's degree by Research, please complete section (a) below. For all other research degrees, please complete both sections (a) and (b) below:*

(a) Hard Copy

I hereby deposit a hard copy of my thesis in the University Library to be made publicly available to readers (please delete as appropriate) ~~EITHER~~ immediately ~~OR after an embargo period of …….. months/years as agreed by the Chair of the Board of Graduate Studies~~.

I agree that my thesis may be photocopied. YES / ~~NO~~

(b) Digital Copy

I hereby deposit a digital copy of my thesis to be held in WRAP and made available via EThOS.

Please choose one of the following options:

~~EITHER~~ My thesis can be made publicly available online. YES / ~~NO~~

~~OR My thesis can be made publicly available only after…..[date]~~
~~YES / NO (*Please delete as appropriate*)~~

~~OR My full thesis cannot be made publicly available online but I am submitting a separately identified additional, abridged version that can be made available online.~~
~~YES / NO (*Please delete as appropriate*)~~

~~OR My thesis cannot be made publicly available online. YES / NO~~

3. **GRANTING OF NON-EXCLUSIVE RIGHTS**

Whether I deposit my Work personally or through an assistant or other agent, I agree to the following:

Rights granted to the University of Warwick and the British Library and the user of the thesis through this agreement are non-exclusive. I retain all rights in the thesis in its present version or future versions. I agree that the institutional repository administrators and the British Library or their agents may, without changing content, digitise and migrate the thesis to any medium or format for the purpose of future preservation and accessibility.

4. **DECLARATIONS**

(a) I DECLARE THAT:

- I am the author and owner of the copyright in the thesis and/or I have the authority of the authors and owners of the copyright in the thesis to make this agreement. Reproduction of any part of this thesis for teaching or in academic or other forms of publication is subject to the normal limitations on the use of copyrighted materials and to the proper and full acknowledgement of its source.

- The digital version of the thesis I am supplying is the same version as the final, hard-bound copy submitted in completion of my degree, once any minor corrections have been completed.

- I have exercised reasonable care to ensure that the thesis is original, and does not to the best of my knowledge break any UK law or other Intellectual Property Right, or contain any confidential material.

- I understand that, through the medium of the Internet, files will be available to automated agents, and may be searched and copied by, for example, text mining and plagiarism detection software.

(b) IF I HAVE AGREED (in Section 2 above) TO MAKE MY THESIS PUBLICLY AVAILABLE DIGITALLY, I ALSO DECLARE THAT:

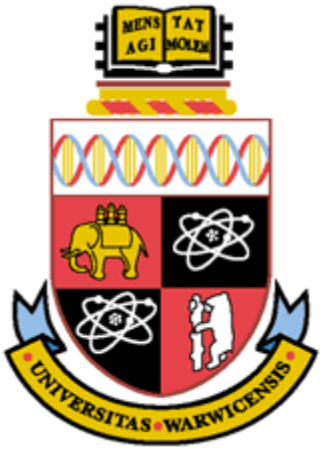
- I grant the University of Warwick and the British Library a licence to make available on the Internet the thesis in digitised format through the Institutional Repository and through the British Library via the EThOS service.

- If my thesis does include any substantial subsidiary material owned by third-party copyright holders, I have sought and obtained permission to include it in any version of my thesis available in digital format and that this permission encompasses the rights that I have granted to the University of Warwick and to the British Library.

5. **LEGAL INFRINGEMENTS**

I understand that neither the University of Warwick nor the British Library have any obligation to take legal action on behalf of myself, or other rights holders, in the event of infringement of intellectual property rights, breach of contract or of any other right, in the thesis.

Student's signature:  Date: April 13, 2015



An Investigation into Chinese Students' Experiences of and Attitudes Towards Tutor Feedback at a Chinese University

by

Christopher James Cookson

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

Doctor of Education

Centre for Education Studies

April 2015

THE UNIVERSITY OF WARWICK

I dedicate this thesis to you, Mum.

Thank you for giving me life – the most precious gift of all.

May you rest in peace and find the happiness in the hereafter you couldn't in life.

I love you and I miss you.

Table of Contents

List of Figures

List of Tables

Acknowledgements

It was only because of the then-still-subsidised tuition rate that I was able, financially, to attend a university in England. I would, therefore, like to devote these first words to expressing my gratitude to the Government of the United Kingdom of Great Britain and Northern Ireland and to its tax-paying citizenry for enabling me to enrol in my course and study for a doctorate degree.

I am also and particularly indebted to my supervisor, Dr. Val Brooks, for her guidance, dedication, professionalism and labours in connection with this thesis but also throughout my attendance at the University of Warwick. I could hardly have accomplished such a great undertaking and achieved the requisite quality for work at this level without her unwavering support and the copious, high-quality and judiciously-given feedback she provided.

My heartfelt thanks also go to my friends and family, particularly my grandparents, for their support during an emotionally difficult time. I would also like to thank all the students involved in my study, the host institution for authorising me to conduct it on their grounds and with their students, and my colleagues for assisting with the dissemination of the questionnaire.

Finally, I am thankful for the opportunities and freedoms that I have had and that I was born into. I am indeed privileged, and I hereby pledge to apply the qualification and the knowledge that I have gained to the betterment of not only myself but also of my fellow people.

Declaration

I hereby declare that the material contained in this thesis is my own work, and that neither this thesis nor any of its contents have previously been submitted for a degree at the University of Warwick or any another university.

Abstract

This thesis explores formative assessment and feedback in China. The study examined how Chinese students at a university in China experienced different feedback types from their Chinese instructors, and what their attitudes towards those were. A mixed-methods approach, using a self-administered, electronically distributed questionnaire and a semi-structured, one-on-one, face-to-face interview, was employed to this end. A total of 232 students – mostly undergraduates – took part, each of whom completed the questionnaire and 30 of whom were interviewed.

A large number of students conveyed an awareness of the formative potential of feedback as well as an expectation towards it playing this role, primarily in terms of learning improvement and being psychologically supportive. Marks and grades were shown to play a dominating and not always unsupportive role. The majority of participants reported overall positive effects of the feedback they were given on aspects of their learning and psychology. However, many felt the feedback they received to be qualitatively as well as quantitatively insufficient.

The study aimed to shed light on a research area and geographical context where accounts in English from a Western perspective are limited. The overarching conclusion of the study's findings is that there appears to be significant convergence between scholars' and Chinese students' expectations of effective feedback. A key recommendation for future research is to conduct further studies into students' perceptions of feedback in non-Western contexts.

Chapter 1

Introduction

1.1 Genesis of the Study

Never previously having received any formal training in educational assessment, I was, when I first started teaching, insufficiently aware of how significant its impact on learners could be and, perhaps even more importantly, of its for-learning, that is to say, instructional or formative potential. The same applies to feedback, a key ingredient of educational and, in particular, formative assessment. Feedback manifested itself for me as an iterative and multidirectional process that could and, indeed, should be used to improve both learning and teaching.

As I honed my assessment skills as a teacher, I also realised how important metafeedback, i.e. students' feedback about the feedback I gave them, was. How could I know whether the feedback I was giving them was having a positive or perhaps a detrimental effect? What was I doing well and what could I improve on? The best way to answer these questions was to ask the students.

At that time, I was in China, a country that had long fascinated me, first from a socio-political perspective and later also culturally and educationally. I felt that formative assessment and feedback – at least how I understood them – faced greater challenges in Chinese classrooms than in Western educational systems due to a large number of factors. I wanted to gain a broader understanding of Chinese students' experiences of and views towards feedback. Thus, I decided to broaden my investigation to all Chinese

students (whose English language skills were sufficiently proficient) at the university where I taught while limiting it to the feedback they received from their Chinese teachers.

1.2 Why Assessment and Feedback?

Assessment is omnipresent: it permeates every facet of modern civilization, and it lies at the heart of socio-economic stratification and a stable, enduring society. According to the meritocratic principle, wealth, income and social status are awarded on the basis of merit, i.e. talent or intelligence, credentials, effort and the like. This, in turn, requires some way of measuring merit. Without assessment, it can be argued that there is no scientific way of testing for competency or conducting a selection process.

A number of voices from the literature seem to agree with this position. Rowntree (1987) wrote: “Assessment permeates every aspect of our lives, and is a natural and automatic activity” (p. 4). Similarly, Stobart (2008) characterised assessment as “a powerful activity which shapes how societies, groups and individuals see themselves” (p. 1). Thus, as Kellough & Kellough (1999) put it, “it is impossible to overemphasize its importance” (p. 419). This, of course, applies to assessment in education as much as it does to that in any other sector. According to Beaumont *et al*. (2011), “assessment has long been viewed as the catalyst for improvement in teaching and learning” (p. 671). Wiliam (2006a) even argued that it is the “only way” by which teachers can ascertain whether their students have learned what they have taught (p. 15). This fact alone lends it a crucial and irrevocable role within education, and places it as a worthy area of research.

Feedback is a consequence of educational assessment (Taras, 2013) and “arguably the most important part of the assessment process” (Price *et al*., 2010, p. 277). It is a key driving force behind giving educational assessment a capacity not only to measure but

also, and perhaps more importantly, to instigate learning. Hattie & Timperley (2007) argued that one important function that assessment tasks can and should perform is to provide feedback in the form of "information and interpretations about the discrepancy between current status and the learning goals" (p. 101). The authors also underscored the importance of both teachers and students seeking and learning from feedback, and that "only when assessment provides such learning is it of value to either" (p. 104). Put another way: only when an assessment is used in conjunction with feedback – be it furnished along with it or sought through it – can it have any learning value.

1.3 The Research Context

The People's Republic of China was founded in 1949. At that time, it was known as "The Sick Man of Asia" because it could not even feed its own people (Jowett, 1991). Since then, China has propelled itself onto the world stage with hitherto unparalleled ferocity and speed. Its growth and influence in the international arena in political, economic, cultural and technological terms is staggering.

In 2001, it acceded to the World Trade Organisation; in 2003, it celebrated its first manned space flight; in 2008, its capital, Beijing, hosted the Summer Olympic Games, shortly before which the *Independent* newspaper had dubbed it a "new superpower" (Milmo, 2008); in the same year, it performed its first spacewalk, becoming one of only nine countries to do so; and, in 2010, Shanghai, China's largest city, hosted the World Exposition. Both Shanghai and Beijing also hosted the APEC summit, in 2001 and 2014, respectively.

Its economic prowess is even more striking: "From 2007 to 2011 China accounted for as much of global economic growth as the G7 leading industrial countries combined"

(Walker, 2011, para. 4) and now has the world's largest Gross Domestic Product (GDP) in terms of purchasing-power-parity (International Monetary Fund, 2014). What is more, China is the world's largest holder of foreign exchange and gold reserves, currently totalling nearly 4 trillion U.S. Dollars (Trading Economics, 2014) – more than one and a half times the U.K.'s entire GDP (Central Intelligence Agency, 2014).

China, however, does not enjoy the same supremacy in terms of education. Only 24% of its population of the relevant age are in tertiary education; this compares to, for example, 59% in both the U.K. and Japan, and 89% in the U.S.A. (United Nations Educational, Scientific and Cultural Organisation Institute of Statistics, 2011). It is a land of extremes, particularly demographically; the pressure on the education system as well as on the job market is severe (Eimer, 2011). The leading cause of death amongst Chinese youths is suicide and the 2014 Report on China's Education – the so-called "Blue Book of Education" – reported that many of these deaths are due to the intense pressure of the exam-oriented education system (Chelala, 2014). Class sizes in China are also infamously large, with 50 being the national average (Organisation for Economic Cooperation and Development, 2010). Perhaps the only positive note in terms of educational achievement on an international level is Shanghai having ranked first in the 2009 and 2012 Program for International Student Assessment test scores in all three categories.

Despite massive economic growth and many larger cities in China having the appearance of modern, Western-style metropolises, the ancient Confucian style of pedagogy is still prevalent (Wang, 2008). According to Confucian thinking, the rigorous acquisition of knowledge is one of the central tenets of learning and personal development. This contrasts distinctly with Socratic thinking prevalent in the West where

the focus is placed on the generation of hypotheses and new knowledge through the private and public questioning and evaluation of widely accepted ideas (Tweed & Lehman, 2002).

Chinese civilisation dates back at least four millennia (Cheng, 2008). For over 2,000 years, it has had an examination system (Xiao & Carless, 2013), beginning as early as the 3rd century BC (Zhan & Wan, 2010) with the *kējǔ* (instituted formally in the 7th century AD) which required applicants to, *inter alia*, understand as well as memorise the Four Books and the Five Classics – the representative works of Confucianism (Berry, 2011). These exams, through which candidates were chosen for civil service positions (Yu & Suen, 2005), are generally regarded as the first standardised tests based on merit (Cheng, 2008). They provided the primary means for social mobility and gave rise to an "examination culture" (Zhan & Wan, 2010). Characterising the *kējǔ*, Han & Yang (2001), explained that:

> First, it strengthened the utilitarian values of education, namely the idea that education to become a government official was the final goal of schooling. Secondly, it stressed the key role in education that was played by examinations. Thirdly, it stressed book knowledge at the expense of practical ability. Fourthly, it laid emphasis on final assessment, neglecting formative assessment (p. 5).

For only just over 100 years has China had a schooling and examination system resembling that in the West. In 1906, a three-tier national examination system took the place of the imperial examinations according to which students were assessed at the conclusion of the three key stages of schooling – primary, and junior and senior secondary (Berry, 2011). However, despite this major overhaul, Berry pointed out that "Examinations were still used as a tool for driving learning and for making summative

judgments of learning" (p. 51). Considering the considerable lifespan of the imperial examination system, this is, perhaps, not surprising.

In 1952, the *gāokǎo* – in English, the National Higher Educational Entrance Examination (also referred to as the National College Entrance Examination) – was implemented to place graduating senior middle school students into a four-level hierarchy of colleges and universities (Yu & Suen, 2005). Ranking highest are the so-called National Key Universities (also referred to as Project 211 universities) as well as Project 985 universities which have been selected to be developed into "world-class universities," of which there are currently 112 and 39, respectively (China Education Center, 2014).

This number represents only a fraction of the total number of institutions of higher learning but, as Yu & Suen (2005) pointed out, enrolment into and graduation from a National Key University is critical for upward social and economic mobility. The National Higher Educational Entrance Examination was abolished during a ten-year period during the Cultural Revolution but was restored in 1976 (Wang, 1996) shortly before Deng Xiaoping initiated sweeping economic reforms under the Open Door Policy.

Still today, the assessment culture continues to be strongly influenced by the system of education and assessment that had existed for millennia (Carless & Lam, 2014). As Dello-Iacovo (2009) observed, although a large number of attempts to reform the education system have been made, "the examination-oriented 'regular system' has proven remarkably resilient, bouncing back with renewed vigour after each assault" (p. 242). She added that Chinese educators have been obdurate in moving away from the examination-based system and that, "left to their own devices, the features of the regular system with its links back to the imperial examination system would inevitably re-

remerge [*sic*]" (*ibid*). Indeed, as she further pointed out, the teaching content still revolves around the exams and most learning consists of rote memorisation. These problems seem to prevail throughout Confucian-heritage countries. Carless (2011b) pointed out that:

> Formative assessment has been shown to be a highly promising strategy for improving student learning […] yet successful implementation in Confucian-heritage (and other) settings is far from easy in view of the dominance of the summative paradigm (p. 51).

What, then, could be behind this stagnation in the Chinese education system and assessment practices? There is immense pressure on all the stake-holders – the schools, the teachers and, in particular, the students. China's enormous population along with the lack of education funding may be considered as the root causes of the problem. Since 1985, the local governments rather than the Chinese Central Government have had to shoulder the majority of the cost of financing compulsory education and the full cost of senior secondary education, leading to funding shortages (Dello-Iacovo, 2009). This has given rise to a policy of establishing not only "key universities" but also "key secondary schools" which enjoy better teaching and learning conditions (Han & Yang, 2001, p. 7).

This policy, in turn, has led to a domino effect. Schools, wanting to send as many of their students to National Key Universities as possible, focus on helping them pass the selective examinations, thereby placing heavy pressure on the teachers. Han & Yang explained that, overwhelmed by large class sizes, teachers are forced to "pay closest attention to the 'best' students, those that have the best prospect of entering university, while tending to neglect the students with lower test achievement" (p. 8). Since teachers' performance is linked to the results of their students' examinations, they place the

greatest focus on an examination-oriented style of pedagogy – a phenomenon sometimes colloquially referred to in the literature as “teaching to the test” (Black & Wiliam, 1998a; Popham, 2001; Torrance, 2012). This typically results in the curriculum being shaped around desired assessment outcomes or to “the assessment tail wagging the curriculum dog” (Barnes, Clarke & Stephens, 2000, p. 624).

Teachers in China also hold the monopoly over the feedback that is provided to students (Chen *et al*., 2014). In China, as in many Confucian-heritage countries, teachers are key figures in society and wield significant power, sometimes challenging or even overriding that of the father (Sully de Luque & Sommer, 2000). In these countries, “culture has somewhat assigned the classroom to be the teacher’s kingdom of sanctity” (p. 407). Learners in these cultures have grown accustomed to a didactic approach where “The stream of knowledge passes along a one-way street from teacher to students” (*ibid*). Chen *et al*. (2012) explained that “The teacher is positioned as the only credible judge or assessor of learning, while students have little sanction to judge or assess each other’s work” (p. 5).

Li & Hui (2007) posited that many teachers in China continue to cling onto more traditional, summative methods of assessment because of their being an established method for producing good results and that good results improve teachers’ image. Consequently, teachers are reluctant to leave their comfort zone and become “pioneers in alternative strategies of assessment” (p. 189). Considering the high-stakes nature of the Chinese education system for both students and teachers, this stance is not surprising. Calling attention to this dilemma, the authors wrote:

> On the one hand, teachers know that the present practices are detrimental to their students' learning, but on the other hand, the cost to bring about innovative assessment strategies is too great because both the teachers and the students cannot afford to perform poorly in competitive scholastic achievement tests which emphasize rote learning (*ibid*).

In this way, as maintained by O'Leary & Brooks (2014), "high-stakes summative assessment has the potential to corrupt the delicately poised relationship between teaching, learning and formative assessment" (p. 532).

Many of the characteristics typical of the Chinese education system described here apply to the setting where the research for this study was conducted. The researcher's time spent working there and several extended, candid conversations with the students about their learning-related experiences confirmed this. Classes typically contain at least 40 students while many hold several hundred, and competition between students is fierce due, primarily, to the great demand for limited number of postgraduate study places, scholarships, work-study placements and other opportunities. Many students define their learning and achievement potential in accordance with their ranking out of dozens or hundreds of others in their department. The pressure emanating from this environment manifests itself noticeably in their psychology as well as learning behaviour.

Although these characteristics are not unique to the Chinese education system, China contrasts both culturally and educationally quite strikingly with American and European contexts. This was one of the major motives for conducting this study on feedback in China. Hattie & Timperley (2007) stated that "feedback is not only differentially given but also differentially received" (p. 100). Similarly, Carless (2006) argued that "One person's helpful suggestion may be another person's judgemental criticism" (p. 229). It is likewise argued here that feedback is not an isolated, homogenous phenomenon; it comes in

many guises, and may impact its recipients in a number of ways and to different extents. This may be all the more the case in a system of assessment as old and as heavily steeped in culture and history as China's.

1.4 Organisation of the Thesis

This chapter has outlined the background, motivation and context behind the study. Chapter two provides a review of the extant literature on feedback as well as in the related area of formative assessment. Chapter three offers a discussion of a variety of relevant methodological aspects, including a consideration and justification of the chosen data collection tools. Chapter four presents, synthesises and interprets the study's findings. Chapter five concludes the thesis with a summary of the results, implications of the findings, limitations of the study, and recommendations for future research.

Chapter 2

Literature Review

2.1 Introduction

During the past half a century and particularly throughout the last two decades, a vast amount of literature on educational assessment has been published, too much for even a doctoral thesis to encompass. An all-text search on Education Research Complete for the phrase “educational assessment” generated 28,096 results while a search for the words “assessment” or “evaluation” in combination with “feedback” and “education” produced 60,490 and 57,500 results, respectively. It can, therefore, be deduced that the field of educational assessment and the core concepts under investigation in this thesis are ones which have attracted considerable attention.

Both empirical and theoretical works, mostly small-scale qualitative and some large-scale quantitative studies are represented, though the vast majority of these were conducted in American and British contexts; the lead authors, likewise, are usually American or British. This is corroborated by a meta-review of assessment research in higher education (Evans, 2013) in which 75% of the 460 included articles were written by lead authors from either Australia, the U.K. or the U.S.A. Moreover, while a reasonable amount of English literature on the Chinese education system in general exists, most research specifically into educational assessment has been conducted in the Anglophone world, particularly in Western Europe (Carless, 2011a; Chen *et al.*, 2012). The vast majority of literature on educational assessment in the Chinese context that does exist appears to be in Chinese. The second two points were two of the driving motivations for the geographical focus of the present study.

While feedback represents the focus of the study, much attention is also given to formative assessment in the literature review that follows. This is because feedback is a fundamental component of formative assessment (Sadler, 1989; Gipps 2012). Black & Wiliam (1998a) also noted that these two concepts "overlap strongly" and that the latter is central in the former (p. 50). In his definition of formative evaluation, Bloom (1969) made explicit mention of feedback as did Black & Wiliam (1998a,b) and Black *et al*. (2003) in their conceptualisation of formative assessment. Hattie & Timperley (2007) tied assessment directly to feedback: "Assessment can be considered to be activities that provide teachers and/or students with feedback information" (p. 101). Taras (2013) even argued that these two concepts have become "increasingly synonymous, […] whether implicitly or explicitly" (p. 33).

As the following review shows, there is a general consensus that feedback and formative assessment in general *can* lead to improvements within education, but a coherent and widely accepted theory which can explain the conditions and mechanisms as well as predict the outcomes of these assessment practices does not yet exist (Torrance, 2012). A commonly reached conclusion at the end of articles on educational assessment is that more research and theoretical exploration need to be undertaken (e.g., Jonsson, 2012; Evans, 2013; Gamlem & Smith, 2013; Nicol, Thomson & Breslin, 2014).

Indeed, this appears to apply specifically in the case of students' perceptions about feedback. Brown (2007), founding his conclusion on the basis of several other scholars, wrote that "research into the student perception of the benefit they receive from assessment feedback is virtually non-existent" (p. 33). Similarly, Poulos & Mahony (2008), who argued that "How the student interprets and deals with feedback is critical to the success of formative assessment" (p. 144), noted that "research on students'

perceptions of feedback remains thin" (*ibid*). Masson (2011) also maintained that "research looking into students' beliefs about feedback is still lacking" compared to teachers' perspectives (p. 189). Writing in relation to the effectiveness of different types of feedback in the Chinese context, Wang & Wu (2012) stated that "perceptions and attitudes of both teachers and students have unfortunately been neglected to a large extent" (p. 288). This reality jars with Yorke's (2003) statement that "The importance of the student's reception of feedback cannot be overstated" (p. 488).

The present study endeavours to make a contribution to knowledge on this subject. Universities, where high numbers of Chinese students are enrolled, and university instructors who come into contact with these students may also benefit from the study's findings. This is the case, for instance, with universities in the U.K. where Chinese nationals represent, by far, the largest group of international students (UK Council for International Student Affairs, 2014); according to the Higher Education Funding Council for England (2014), there are nearly as many Chinese as British students entering full-time, taught postgraduate courses.

2.2 Organisation of the Chapter

After providing a brief explanation of the methodology employed to search for and select literature, the chapter demonstrates how formative assessment evolved as a concept and how it was promoted as a mechanism to support learning. The next section critically considers the term "formative" and makes a case for a particular conceptualisation. The subsequent section discusses some of the most important contributions made within the discussion on feedback. Open questions related to feedback in higher education are considered in the section that follows. The next section examines a specific type of feedback which represents the phenomenological focus of the study, and briefly argues

for the usage of another term which more accurately embodies this focus conceptually. The final major section gives insights into educational assessment in the Chinese milieu, first by considering the Chinese Central Government's stance on and its implementation specifically of formative assessment before reviewing the available literature in English on formative assessment and feedback in China. The chapter concludes with some conclusions and outlines the study's research questions.

2.3 Literature Search Methodology

All articles included in the literature review as well as other parts of this thesis were searched and accessed electronically through several major online academic databases, including Education Research Complete, the Education Resources Information Center (ERIC), Journal Storage (JSTOR), ScienceDirect and Wiley Online Library. Searches were also performed through the University of Warwick Library online catalogue as well as directly through websites of scholarly publishers, including Taylor & Francis Online, SAGE Journals and SpringerLink. A number of sources were also accessed through Google Scholar and Google Books. The search was limited to articles in English and containing various permutations and inflections of the keywords "assessment," "evaluation," "feedback" and "formative." The search for articles related to China was narrowed by additional usage of the keywords "China" and "Chinese." The search for these articles was also restricted to those published in or after 2001, principally since it was in this year that China began to reform its education system.

Each search typically produced several dozens or hundreds of results. In cases of higher numbers of results, searches were, where possible, limited to titles and abstracts. Those articles which were identified as seminal in the field of educational assessment as well as those which appeared to be particularly relevant to the study's research questions

were given priority. A valuable tool for ascertaining the prominence of particular works and their candidacy for inclusion in the literature review was Google Scholar which can be used to determine the approximate number of citations. The primary method for doing this, however, was to manually consult and compare a number of more current meta-analyses and other works of high relevance; frequently cited works were, then, searched individually. This method was continued with successive articles in order to discover additional works not found through earlier searches.

A number of challenges presented themselves regarding finding literature from Chinese scholars on feedback and even educational assessment more generally. As indicated in Section 2.1, one was a relatively scarce amount of context-specific literature in English. Another problem, which is faced by even those who are able to read Chinese, is that articles written by Chinese or China-based scholars are not widely accessible to the international academic community at large. One particularly comprehensive resource does exist: the China Knowledge Resource Integrated Database which covers millions of Chinese publications on a variety of subjects in full-text format which may be downloaded with a subscription. However, despite the website offering an English interface, few of the articles appear to be in English. Consequently, the literature reviewed here likely represents only a small fraction of the extant literature on this particular subject.

A further problem was that the majority of those publications available in English appeared to concern investigations conducted outside of mainland China, particularly in Hong Kong where, after British rule lasting 156 years, a different cultural environment and educational system predominate (Cheung & Hui, 2003; Ko, 2012); Carless (2006) also noted that Hong Kong Chinese students have “particular characteristics,” though did

not elaborate on this point. These studies cannot, therefore, be included here, which further diminishes the amount of suitable literature. In addition, the majority of studies limit their focus to English language learning and give little attention explicitly to feedback which is not differentiated from the broader process of formative assessment. Consequently, very little could be found in terms of studies in China and with Chinese participants specifically on teacher feedback received by students.

The goal of this particular segment of the literature review is, thus, to provide an overview of the employment of and attitudes toward formative assessment and feedback in classrooms in mainland China which is as complete as the circumstances allow. The name “mainland China” is understood here to mean only those territories which are under the direct and full jurisdiction of the Chinese Central Government, thereby excluding Hong Kong, Macao and Taiwan. Since education policy in these three territories is not determined by the Chinese Central Government, any official guidelines for or any studies and reports concerning formative assessment originating there can, therefore, not be representative of the rest of China and do not receive mention in this chapter. For the sake of simplicity, the word “mainland” is, however, normally not used.

2.4 Formative Assessment

The idea of using data gathered from educational assessments for formative purposes goes back at least seven decades. Writing on the subject as early as 1942, Tyler highlighted the importance of using evaluation data, as he termed it, for continuous (curriculum) improvement (Haertel & Herman, 2005). It took another quarter of a century before Scriven (1967), an American writer, coined and systematically distinguished between the terms “formative evaluation” and “summative evaluation” [Note: the term “evaluation” is more commonly used by American writers (Sadler, 1989) who, in contrast

to their British counterparts, do so in reference to not only the appraisal of programmes or organisations, but also of individuals (James, 2000)].

Scriven (1967) differentiated between the goals and roles of evaluation, and deemed formative evaluation as a tool for collecting information to evaluate the effectiveness of a particular curriculum. That information could, then, be used to make improvements to the curriculum or to inform the adoption of a new curriculum. However, Scriven spoke cautiously of the "intrusion" of evaluation into education and that of evaluators into the curriculum-making process, criticising what he referred to as "testing for learning" as being potentially "destructive" and "incompetent" (p. 9). Building on the work of Atkin (1963), one argument he put forth in justifying this view was that it may interrupt the natural and proper learning process at too early a stage.

Despite this, Scriven appears to have captured the essence of formative assessment, although his understanding of the concept was directed at programme assessment rather than student assessment. Even years later, this was still apparent:

> Formative evaluation […] is typically conducted *during* the development or improvement of a program […] and it is conducted, often more than once, *for* the in-house staff of the program with the intent to improve (emphasis in original, 1991, pp. 168-169).

Another point worth noting in the above quotation is that, by emphasising the word "during," Scriven placed importance on *when* the evaluation took place. This point is reinforced by his statement on the "summative evaluation of a program [which] […] is conducted *after* completion of the program […] and for the benefit of some external audience or decision-maker […]" (emphasis added, p. 340).

In contrast to Scriven, Bloom (1969), another American scholar, placed more importance on the purpose of assessment when characterising formative evaluation. He stated:

> 'formative evaluation' [is used] to provide feedback and correctives at each stage in the teaching-learning process. By formative evaluation we mean evaluation by brief tests used by teachers and students as aids in the learning process. While such tests may be graded and used as part of the judging and classificatory function of evaluation, we see much more effective use of formative evaluation if it is separated from the grading process and used primarily as an aid to teaching (p. 48).

Thus, Bloom held that assessments were neither summative nor formative *per se*. Rather, any assessment could take on either function. Looking at the above quotation, it does appear, however, that he argued for using formative evaluation to support instruction rather than scoring. In fact, by considering the word "aid," one can acquire a first taste of what Bloom really foresaw for formative evaluation: to *improve* the learning and teaching processes. Writing two years later, he and his co-authors stated:

> Formative evaluation is for us the use of systematic evaluation in the process of curriculum construction, teaching, and learning for the purpose of *improving* any of these three processes (emphasis added, Bloom, Hastings & Madaus, 1971, p. 117).

Regarding the lattermost process, the authors were more specific:

> The purpose [of formative observations] is not to grade or certify the learner; it is to help both the learner and the teacher focus upon the particular learning necessary for *movement towards mastery*" (emphasis added, p. 61).

In mastery learning, "the students are helped to master each learning unit before proceeding to a more advanced learning task" (Bloom & Sosniak, 1985, p. 4). Therefore, improvement – in terms of learning objectives being met on a step-by-step basis – was, for Bloom, an integral element of formative evaluation. In addition to purpose, Bloom,

Hastings & Madaus (1971) also identified timing and the level of generalisation as key factors in delineating formative evaluation.

Despite placing the emphasis on different characteristics, there was significant convergence between Scriven and Bloom in that they both argued that formative evaluation should not only change something, but that it should also be employed during the teaching and learning process rather than after it. Some years later, Sadler (1989), an Australian academic, deepened the discussion. In his conception, the key attributes of what he preferred to call "formative assessment" should be related to purpose and effect, not to timing (p. 120). According to him,

> Formative assessment is concerned with how judgements about the quality of student responses (performances, pieces, or works) can be used to *shape and improve* the student's competence by short-circuiting the randomness and inefficiency of trial-and-error learning (emphasis added, *ibid*).

The essence is, as Wiliam (2006b) also pointed out, the same as that foreseen by both Scriven and Bloom: to change and improve teaching and learning. Sadler (1989) also made it clear that "evaluation as curriculum content should be clearly distinguished from evaluation as an agent," the latter of which "is inextricably connected with constructive activity, and is primarily enabling and facilitative rather than an end in itself" (p. 138). Sadler emphasised that formative or, as he put it, facilitative evaluation (or assessment) is that which has as its subject the students and the work they produce. Indeed, Bennett (2011) affirmed that, even though Bloom's understanding of the concept still essentially holds today, formative assessment is now used in reference to students rather than programmes.

Sadler (1989) highlighted, furthermore, the importance of collaboration between students and their teachers, and of that between students themselves in the formative assessment process. He underscored that students should be formatively assessed not only by their teacher but also by themselves – a practice he called "self-monitoring." According to this idea, students are provided with a particular learning target by the teacher which, ideally, they will internalise, culminating in setting their own goals and monitoring their progress thereto. In a later work, Sadler (1998), using the terms "self-assessment" and "peer-assessment," added that feedback produced by students for students should be communicated at a level and in a manner which they can understand.

Around this time, using assessment to guide and improve learning and teaching rather than only to measure them became a fully-fledged movement and a new paradigm (Broadfoot, 1993; Stefani, 1998; Buhagiar, 2007; Crossouard, 2009). The emerging view was that "assessment does not stand outside teaching and learning, but stands in dynamic interaction with it" (Gipps, 1994, p. 15). In other words, it was no longer unidirectional or simply an end-product but an iterative, reflective process.

Reflecting on that time, Cross (1998) wrote of "a clear shift from assessment for accountability to assessment for improvement" (p. 5) where students were no longer seen as only the subjects of assessment but also as "active and intelligent" participants in it (p. 6). She argued that student learning could not be improved without the participation of the students themselves; they, as lifelong learners, must become more independent and take on greater responsibility.

To facilitate Cross' idea, Elwood & Klenowski (2002) supported the creation of so-called "communities of shared practice" [see also Wiliam (1998) for a similar exposition on

"communities of practice"] in which the students are seen as "participants in a learning curriculum where understandings concerning what they are doing, or, in relation to assessment, what is being done to them are shared" (p. 246). In these communities, moreover, students also engage in peer-assessment and use the feedback they receive from their peers to improve their learning.

In the U.K., where the paradigmatic shift was particularly strong, formative assessment was formalised as "Assessment for Learning" – a term the coining of which is often attributed to Gipps (1994) but which, as Stobart (2008) and Wiliam (2012) pointed out, was also being used independently at that time by James (1992) and Sutton (1995). At least within the compulsory education sector, Assessment for Learning was given the aim of focusing on classroom learning (Taras, 2008). On the organisational level, this movement was spearheaded by the Assessment Reform Group which formally disassociated Assessment for Learning from its summative cousin, "assessment *of* learning," and conceived it as "the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there" (emphasis added, 2002, pp. 1-2).

This interpretation was formally adopted by the Department for Education and Skills, the Qualifications and Curriculum Authority, and the Office of Standards in Education (Looney & Wiliam, 2005). Subsequently, U.K. Minister of State for Schools and 14-19 Learners announced in a twenty-one page document entitled *The Assessment for Learning Strategy* that the Government would promote Assessment for Learning by investing £150 million in the continuing professional development for teachers in Assessment for Learning over the following three years (Department for Children, Schools and Families, 2008).

Numerous scholars have attested to the potential of formative assessment to improve both learning and teaching (e.g., Fuchs & Fuchs, 1986; Crooks, 1988; Kluger & DeNisi, 1996; Gipps, 1999; Harlen & Deakin Crick, 2003; Brookhart, 2007; Hattie & Timperley, 2007; Shute, 2008). At the forefront of the new wave of support for formative assessment and Assessment for Learning are Black and Wiliam, both British scholars. It is worth pointing out that nearly all the articles on formative assessment in China reviewed in Section 2.9 use Black and Wiliam's interpretation of formative assessment, despite an occasional preference for the term "evaluation" or "appraisal" in lieu of "assessment."

The year 1998 saw two highly influential publications from these authors. Their paper entitled *Assessment and Classroom Learning* (1998a) was the product of an extensive review and meta-analysis of over 250 articles related to formative assessment. Reporting effect sizes of one half to a full standard deviation, the authors pronounced that their synthesis of research "shows conclusively that formative assessment does improve learning" (p. 61). They also found that formative assessment "helps low achievers more than other students and so reduces the range of achievement while raising achievement overall" (p. 141).

It is worth pointing out, however, that the claims made by Black & Wiliam and other proponents regarding the effectiveness of formative assessment in terms of student learning improvement have been questioned by a number of writers, including Perrenoud (1998), Smith & Gorard (2005), Dunn & Mulvenon (2009) and Bennett (2011). Biggs (1998), who took issue primarily with Black & Wiliam's (1998a) short-sighted approach to assessment – in terms of the necessarily negative impact of summative assessment and their exclusion of the same –, also criticised their review for being

"loose and sometimes repetitive" and with "some arguable content issues" (p. 105). Black *et al*. (2005) also acknowledged that formative assessment may not result in improved student achievement "in *all* classes, with *all* teachers on *all* occasions," and reduced their claim to formative assessment being an effective intervention in general (emphasis in original, p. 7). The point of student achievement in terms of its relationship with feedback is returned to in Section 2.6.

Their second publication, *Inside the Black Box: Raising standards through classroom assessment* (1998b), for which the previous article acted as the source work, received even greater attention. As did Sadler (1989) before them, Black & Wiliam (1998b) emphasised heavily the centrality of dialog between students and teachers, stating that it "should be thoughtful, reflective, focused to evoke and explore understanding, and conducted so that all pupils have an opportunity to think and to express their ideas" (p. 12). Likewise, they stressed the essential property of formative assessment as lying in the evidence from assessments being used by both learners and teachers to regulate or adapt learning and teaching, which is reflected by their definition of this concept:

> We use the general term *assessment* to refer to all those activities undertaken by teachers – and by their students in assessing themselves – that provide information to be used as feedback to modify teaching and learning activities. Such assessment becomes *formative assessment* when the evidence is actually used to adapt the teaching to meet student needs (emphasis in original, p. 140).

In addition to modifying the content of a given lesson, other aspects, including learning strategies, teaching methods and the way in which feedback is provided, may, therefore, also be altered. Consequently, formative assessment incorporates not only the perspective of the teacher in terms of the provision of instruction, but also that of the learner in terms of the reception of the material being taught.

In addition to the above characteristics of formative assessment, in an article commissioned by the Assessment Reform Group, Black & Wiliam recognised that one of the "deceptively simple, key factors" for improving learning through assessment was "a recognition of the profound influence assessment has on the motivation and self-esteem of pupils, both of which are crucial influences on learning" (Assessment Reform Group, 1999, p. 4). They added: "Assessment that promotes learning […] is underpinned by confidence that every student can improve" (*ibid*). Thus, it is clear that, when defining assessment that improves learning, Black & Wiliam not only considered the involvement of students in the assessment process and the provision of two-way communication between them and their teachers but also the effects assessment has on their psychology, principally on their motivation and confidence.

For this reason, Black & Wiliam (1998b) suggested that feedback to students should focus on their work and on how they can improve, warning against using it for making normative comparisons with other students. As Wiliam (2005) put it, "one should criticise the behaviour, not the child" (p. 25). This principle was formalised a few years later in a poster developed by the Assessment Reform Group (2002) in which it delineated ten principles of Assessment for Learning, two of which included: "Assessment for learning should be sensitive and constructive because any assessment has an emotional impact" and "Assessment should take account of learner motivation." The point of motivation in terms of its relationship with feedback is returned to in Section 2.6.

2.5 The "Formative" Element

Before progressing to the subject of feedback, it is necessary to pose an important question which has also been raised in the literature: what does the notion "formative" actually mean and which effects on learners does it entail? There appears to be some

confusion on this topic. Writing in reference to Assessment for Learning, Stobart (2008), for example, posited that it "has been introduced as a term partly because of the many misunderstandings that 'formative' generates" (p. 146). Indeed, its meaning can vary depending on whether the focus is on the purpose or the result.

As Sadler (1989) pointed out, the adjective "formative" etymologically denotes forming, shaping or moulding something (or someone), usually to achieve a desired (and presumably positive) end. In a revised definition of formative assessment, Black *et al*. (2003) appear to emphasise the usage of assessment information rather than a particular outcome thereof as being the decisive factor:

> An assessment activity can help learning if it provides information to be used as feedback by teachers, and by their students 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 the teaching work to meet learning needs (emphasis in original, p. 2).

Indeed, the formative use of summative tests was one of the areas developed – albeit unintentionally – during the King's-Medway-Oxfordshire Formative Assessment Project reported by Black *et al*. (2003). Originally, the participating teachers were recommended to avoid summative assessment. However, they found that formative and summative assessment were too closely interrelated to be separated, and so found ways of reconciling the two. Similarly, Biggs (1998) asserted that there is "a powerful interaction" between formative and summative assessment "that could usefully be incorporated in an overall synthesis," enabling the effects of both to be "conceptualised within the same framework" (p. 106). Taras (2002) also argued that "formative assessment, and therefore feedback, is essential both for judging work (either by tutors or students) and for permitting learning to become a logical outcome" (p. 504).

Acknowledging this problem, Black & Wiliam (2009) further modified their definition of formative assessment in which they not only reiterated its conceptual essence but also offered a refined explication of the term “formative:”

> Practice in a classroom is formative 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 taken in the absence of the evidence that was elicited (p. 9).

From this definition, it can be argued that the question of what “formative” means is simply one of semantics and that, etymological and linguistic analysis aside, it – when used in conjunction with assessment – always carries a positive connotation in terms of learning and instruction. The reason for this is decisions taken by learners or their teachers need not *actually* be better, but simply be *better founded*. Hence, even if the end-result of a practice or a decision made by a learner or a teacher is undesirable, it can still be referred to as formative assessment, providing it was well-intended.

Black & Wiliam recognised this quandary and admitted that collecting and using evidence with merely the intention of making better decisions would, according to their definition, “unfortunate[ly]” still be considered as formative (p. 10). Wiliam (2006a) referred to this as formative intention rather than formative action. In fact, even before their seminal publications, Wiliam & Black (1996), wrote that “an assessment that is intended to be formative (i.e. has a formative purpose) but does not, ultimately, have the intended effect (i.e. lacks a formative function), would not, with this definition, be regarded as formative” (pp. 543-544).

As an alternative to the intention, therefore, Black & Wiliam (2009) considered using the outcome as the focus of their definition. They required that any instructional adjustments improve learning to a higher degree than would have been the case in the absence of those adjustments. However, again, they admitted that such a requirement would not be fruitful since learning is unpredictable; even actions with a record of success in terms of producing learning may prove ineffective under certain circumstances or with particular learners. Furthermore, it would not be feasible to ascertain whether the resulting action (or outcome) was more positive than it would otherwise have been in absence of the formative assessment event; it is, as they pointed out, a hypothetical situation involving a counter-factual claim.

Hence, it may be difficult to establish whether a particular assessment event was truly formative as is the meaning of "formative." Black & Wiliam (2009) suggested that assessment be deemed as formative if so-called "moments of contingency" were created and capitalised upon for the purpose of regulating learning processes, whether synchronously or asynchronously (p. 10). The view taken here is that the notion of "formative" may be more appropriately understood as a guiding principle for the improvement of learning and instruction. Furthermore, although purpose may play an important role in facilitating particular outcomes, the key in determining whether an assessment act is or was formative ought to be that the information collected from it was indeed utilised to enhance learning or teaching, even when such efforts were, ultimately, fruitless.

This parallels the position paper of the *Third International Conference on Assessment for Learning* which stated that "Sources of evidence are formative if, and only if, students and teachers use the information they provide to enhance learning" (emphasis in original,

Klenowski, 2009, p. 264). Although this understanding implies the intention to enhance learning, the key word (already emphasised) is the word “use.” In other words, the evidence should not simply be collected and remain unused. Sadler (1989) explained that, “If the information is simply recorded, passed to a third party who lacks either the knowledge or the power to change the outcome, or is too deeply coded […] to lead to appropriate action, the control loop cannot be closed” (p. 121). In other words, unused or unusable assessment information can have no effect and is, therefore, not formative.

If the information gained from the assessment is used, it will, in some way and to some extent, impact those for whom it was intended, thereby also meeting at least the basic semantic requirement of the word “formative” as noted earlier. According to Torrance (2012), “assessment is always formative, it will always impact on students and have a central place in what and how students learn, but not necessarily in a positive sense” (p. 334). The above understanding of the term “formative” still applies even if the impact is not visible.

Leahy & Wiliam (2012) explained that “formative assessment need not alter instruction to be formative – it may simply confirm that the proposed course of action is indeed the most appropriate” (p. 51). To clarify, even though the instruction may not have been altered, this is still an example of formative action since confirmation of the effectiveness of the curriculum in question *(pro)actively* avoided the instatement of another, potentially less effective one. As Black & Wiliam (2009) put it, a particular course of action confirmed in this way through the assessment information elicited, though unaltered, is “better grounded in evidence” (p. 10). In fact, even more positively and more accurately, this would constitute a for-learning event.

2.6 Feedback

Feedback plays a crucial role within education. In the words of Rowntree (1987), "Feedback, or knowledge of results, is the lifeblood of learning" (p. 24). The definition of feedback has evolved over time. Kulhavy (1977) conceived feedback as "[…] any of the numerous procedures that are used to tell a learner if an instructional response is right or wrong" (p. 211). Sadler (2010) clarified that this earlier understanding of feedback – "now viewed as unduly narrow" (p. 535) – implies that, once this feedback has been imparted, learners would go on to revise as appropriate in order to remedy any shortcomings.

In a frequently cited work, Ramaprasad (1983) defined feedback as "information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way" (p. 4). He underscored that this information "is feedback only when it is used to alter the gap" (p. 6). This proviso is crucial since it distinguishes information which is only given and received from that which is also used. As discussed in Section 2.4, Black *et al.* (2003) used the same principle to delineate formative assessment. It would emanate from this that information provided by the teacher – no matter of how high a quality it may be – cannot, strictly, be deemed as feedback if the recipient did not apply it. In other words, true feedback cannot be given; it can only evolve *ex post facto*. Moreover, it rests on both tutors and learners to ensure that this information-to-feedback metamorphosis takes place.

Admittedly, however, common usage of the term may not entail such an exacting conceptualisation; for many, it may simply equate to the comments provided by someone on someone else's work. This is, in fact, how Kluger & DeNisi (1996) – who used the phrase "feedback interventions" (that is, information on task-performance) –

understood it: “actions taken by an external agent to provide information regarding some aspects of one's task performance” (p. 255). Consequently, it may be useful to differentiate between “giving feedback” to reflect the latter process and “feeding back” to indicate the former.

In a similar vein to Ramaprasad’s (1983) reference to the closure of the gap, Sadler (1989) wrote that the defining features of feedback must be that it “requires knowledge of the standard or goal, skills in making multicriterion comparisons, and the development of ways and means for reducing the discrepancy between what is produced and what is aimed for” (p. 142). Therefore, for both Ramaprasad and Sadler, true feedback which improves learning must not only be targeted and inform the learner of the reference level, but it must also be *used* by them.

As noted in Section 2.4, Sadler (1998, 2010) espoused a proficient usage of feedback by the learners themselves rather than simply its provision by their teachers as the ultimate goal of classroom assessment. In fact, he endorsed these assessment practices and teachers’ expertise in them consciously being made part of the curriculum in order for students to become truly independent learners and effective self-assessors.

Hattie & Timperley (2007) defined feedback as “information provided by an agent […] regarding aspects of one’s performance or understanding” and described it as a “‘consequence’ of learning” (p. 81). In order to understand the purpose, effects and types of feedback, the authors suggested imagining a continuum of instruction and feedback:

> when feedback is combined with more a correctional review, the feedback and instruction become intertwined until ‘the process itself takes on the forms of new

> instruction, rather than informing the student solely about correctness' (Kulhavy, 1977, p. 212) (p. 82).

There are several influential works which have highlighted the effects of feedback on various facets of learning. In a meta-review frequently cited first on this subject, Crooks (1988) weighed the results from studies investigating the relationship between classroom evaluation and students' learning strategies, motivation and achievement – all elements which are included in the present study. Although Crooks' review focused primarily on how evaluations and evaluative processes affect students rather than specifically the feedback they receive from or during those evaluations and evaluative processes, he did enter into a brief discussion on the impact of other instructional practices that involve evaluation, in which he included feedback.

He commented that feedback does appear to impact students' performance, but that the nature and size of the effect varies according to a number of factors, including the characterisation and extent of the feedback, its timing, how well it takes the recipient's existing level of performance into account, and how it relates to the summative functions of evaluation. Crooks argued that feedback should be given while it is still relevant, that is, usually "soon after a task is completed" (p. 469).

The moment of relevance he chose for the provision of feedback may be questioned since feedback could be equally or perhaps even more relevant while a task is being carried out, that is, before it is completed. However, it is likely his key point was that this timing allows students to respond to the feedback while the task is still fresh in their memory and while they can still relate to it. This, in turn, would better pre-empt the repetition of errors whose correction becomes more recalcitrant the more deeply they become engrained.

According to Crooks, students should also be given the opportunity to demonstrate that they have modified their learning behaviour on the basis of that feedback. In addition, it should be specific and needs-based while not being overwhelming in quantity or depth. With respect to motivation, Crooks noted that feedback is best tied to students' progression towards mastery rather than to comparisons with their peers, particularly in the case of young, unconfident or poor-performing students. Praising feedback should be given sparingly and be task-specific, and critical comments are best avoided.

There is also a number of individual studies which investigated the cognitive and affective impacts of feedback on students, and which showed that this process occurs differentially according to how feedback is targeted, as will be demonstrated in the following. The most prominent foci in the literature are self-efficacy, self-regulated learning, motivation and achievement. A number of the most important contributions on these sometimes-overlapping aspects will be reviewed in the following.

Bandura, a particularly influential contributor to theory on self-efficacy, understood this construct as the "belief in one's capabilities to organise and execute the course of action required to produce given attainments" (1997, p. 3). In other words, self-efficacy refers to a person's feeling of capability to succeed in a particular situation or with a particular task. In an earlier work, he hypothesised that "expectations of personal efficacy determine whether coping behavior will be initiated, how much effort will be expended, and how long it will be sustained in the face of obstacles and aversive experiences" (1977, p. 191).

Schunk, who collaborated with Bandura on at least one occasion (Bandura & Schunk, 1981) and who used his self-efficacy theory as a conceptual reference point, is one of

the most prolific writers in this area. Through a study involving 40 elementary-level students who were assessed for their perceived self-efficacy for solving subtraction problems, Schunk (1982) determined that effort attributional feedback that linked past achievement with effort promoted, *inter alia*, perceived self-efficacy more so than that which stressed the value of future effort.

In a very similar study in which he compared the effects of ability and effort attributional feedback both on children's perceived self-efficacy and achievement, Schunk (1983) found that those who only received ability feedback demonstrated the highest subtraction skills and self-efficacy; in addition, only effort attributional individual feedback and that in conjunction with ability attributional feedback were shown to outperform the no-feedback condition. In an experiment with 386 Chinese primary and secondary school children in Hong Kong, Hau & Salili (1996) replicated Schunk's (1983) results of Grade 3 students with their Grade 4 students, but found that Grade 10 students, who had not been given ability feedback, performed better on the difficult items compared to those who had been given ability feedback.

Schunk & Rice (1991) cautioned, however, that, although ability feedback becomes more credible as skills develop and success theoretically requires less effort, it may detract from the learning focus of goals. That is, ability feedback may give students the impression that they have a particular skill or the ability to achieve a particular goal, but whether they have progressed towards actually accomplishing that goal or learning task is another matter. Therefore, according to the authors, the stage of task accomplishment needs to be taken into account.

In other articles, Schunk (1984, 1985) held that, in order to support self-efficacy, assessments of task performance should emphasize performance or task mastery (i.e. the quality of learners' work) rather than task engagement (i.e. completion of the task) and social comparison. Moreover, feedback that acts to persuade learners that they possess the capabilities to perform a task augments self-efficacy, but this enhancement will be temporary if successive efforts result in poor performance (Schunk & Rice, 1991). In other studies, Schunk & Swartz (1993) and Schunk (2003) demonstrated that, particularly in situations where learners cannot reliably self-assess, feedback on progress towards a goal that conveys that the strategy is useful and that students are capable of continuing to improve their skills can raise self-efficacy.

In a review of research on the impact of assessment on pupils' motivation, Harlen & Deakin Crick (2002) stated:

> How assessment of their learning is reported back to the pupil (feedback) affects motivation to learn. It has a central role since the feeling of self-efficacy is derived from performance in previous tasks of the same kind. If pupils have experienced success in earlier performance they are more likely to feel able to succeed in a new task (pp. 5-6).

Harlen & Deakin Crick understood motivation as an umbrella concept which encompasses a number of variables, including self-efficacy, sense of self as a learner, self-esteem, self-regulation, interest, effort, goal orientation and locus of control. Similar to Crooks (1988), Harlen & Deakin Crick (2002) asserted that feedback related to the task rather than to the students themselves produces "greater interest and effort," and that feedback which compares students through marks or grades leads to students focusing on achieving higher scores rather than on learning or depth of understanding (p. 6). Feedback should be provided in relation to the learning goals and how to achieve

them. This type of feedback is more amenable to both the learning process and students' motivation.

An earlier but particularly prominent contribution on the effects of feedback on students' motivation is Butler's (1987) study on task-involving *versus* ego-involving feedback. She argued that higher levels of interest and intrinsic motivation stem from greater involvement in the task rather than from focus on the self, which has the opposite effect. She reasoned further that positive feedback on the task supports interest in the same whereas that on the self will decrease it.

Butler classed grades and praise as ego-involving feedback since they focus on the self while regarding personalised comments as task-involving feedback since they focus on the task. She confirmed her assertions through an experiment in which subjects were either given no feedback or one of three types of feedback, including comments, praise or grades before being assessed in terms of interest in continuing the task. The results showed that the group that had received comments had the highest level of interest in carrying on with the activity.

In an influential theoretical piece, Butler & Winne (1995) accentuated students' deeper involvement in the feedback process, asserting that self-regulated learners are amongst the most effective learners and that feedback is the "prime determiner" of the self-regulation process (p. 245). According to the authors, learners who are able to self-regulate are able to control and evaluate their own learning processes. They singled out five functions of feedback, including confirming conceptual understandings, adding to the same, overwriting incorrect knowledge or prior beliefs, further refining or "tuning" understandings, and restructuring theoretical schemata.

Rather than viewing feedback simply within an input-output model, Butler & Winne argued that, once it has entered the cognitive system of a learner, it triggers a series of internal, cyclic processes. These processes produce idiosyncratic interpretations of external feedback (i.e. internal feedback) and differing effect magnitudes of the feedback received that depend on a number of pre-existing learner characteristics, including prior knowledge and beliefs, domain-specific knowledge, strategy knowledge, and motivational beliefs, in particular self-efficacy.

Pivotal in Butler & Winne's model of self-regulated learning is monitoring, especially that of calibration which they term as a "meta-monitoring or self-regulating function" (p. 272). Through self-monitoring, learners continually assess their ability to predict their own understanding. How deeply external feedback is processed is predicated on the degree to which it confirms or disconfirms their confidence in being able to make such predictions. According to the authors,

> learning improves when feedback informs students about their monitoring of learning needs (achievement relative to goals in prior phases of engagement) and guides them in how to achieve learning objectives (cognitive engagement by applying tactics and strategies) (p. 273).

In addition to providing prompts as to which learning strategies are most likely to be effective regarding particular learning tasks or in certain domains, feedback should, therefore, also aid students in becoming more capable self-monitors.

The impact of feedback on student achievement appears disputed. A key earlier review of 40 research studies from Bangert-Drowns *et al*. (1991) on found that, surprisingly, "On average, feedback seems to make a positive but small contribution to achievement," leading to an increase in scores of only one quarter of one standard deviation (p. 224).

They also observed that student achievement was significantly better supported through providing the correct answer than through verification feedback (i.e. right/wrong responses); providing explanations, in turn, produced a greater effect size than only providing the correct answer.

In a landmark review on how feedback affects student achievement, Kluger & DeNisi (1996) examined 131 works selected from over three thousand articles published throughout most of the 20th century. They found that, although feedback improved learning on the whole, it actually worsened learners' performance in almost half of the cases. Based on their meta-analysis of the experimental findings of those works included in their review, they constructed a preliminary theory for investigating the effects of feedback interventions and the role of so-called moderators, including time limitations, task types and task complexity, written and oral feedback, and praise. According to their theory, feedback interventions cause learners' attention to shift between three levels of control: the self, the task in terms of task-motivation processes, and the task in terms of task-learning processes. The authors concluded that feedback directed at the task – simpler ones, in particular – promotes learning and achievement to a greater extent than that directed at the self.

In their paper formulating a conceptual architecture for the design of "psychologically well-founded informative tutoring feedback" within multimedia learning environments, Narciss & Huth (2004) concluded that such feedback which is "systematically designed" not only has positive effects on student achievement but also on motivation. They delineated three primary components of feedback – its content, function and presentation – and two key contributory factors – the learner and the instruction – which they also divided into several elements. Considered within the instructional context are

particular pedagogic objectives, learning tasks and sources of learning problems from typical errors or incorrect strategies. Included in learner-specific information, seen as key to the design of feedback, are students' learning objectives, prior knowledge and skills, and their academic motivation. The authors' overarching argument was that the adaptation of the three primary components of feedback should be guided by these instructional elements and learner characteristics in order to maximise the effectiveness of feedback.

Hattie, another key contributor to the discussion on feedback, has long recognised the centrality of feedback in student learning. He reported that it is the single most powerful influence on student achievement (1987). Even more poignantly, he argued that feedback, served in "dollops," is "The simplest prescription for improving education" (1999, p. 9) and its power is dominant in any list of factors that affect student outcomes (2009).

One of the most frequently cited articles on feedback, *The Power of Feedback* (Hattie & Timperley, 2007), provides a conceptual framework for feedback and assesses the impact – both negative and positive – that it has on learning and achievement. Perhaps the key feature of this publication is its differentiation of the impact according to the type of feedback and the way in which it is provided. Although not adopted directly, this work provided an invaluable blueprint for the formulation and inclusion of several aspects in the present study. Because of this, but also due to its high impact within the literature [Black & Wiliam (2009), for example, devoted an entire section to it] and since feedback represents the core theme of this study, this contribution is reviewed here in greater detail. Those aspects directly relevant to this study are also drawn out.

According to the authors, feedback should "reduce discrepancies between current understandings and performance and a goal" (p. 86). In order for feedback to do this, it must answer three primary questions posed by the tutor or the learner: "Where am I going?," "How am I going?" and "Where to next?" which correspond to the concepts of "feed up," "feed back" and "feed forward," respectively (*ibid*). The first concept refers to the goals which must be clear and attainable. The second concept relates to the progress a student is making towards a task or goal, or towards the successful completion of a part thereof.

The third concept concerns which steps need to be taken in order to make better progress. Hattie & Timperley stated that "This feed-forward question can have some of the most powerful impacts on learning" (p. 90) since, rather than the answer being additional information, tasks and expectations, it should provide "information that leads to greater possibilities for learning" (*ibid*). One of the applications of this information the authors mentioned is "more strategies and processes to work on the tasks" (*ibid*).

By looking at this particular "possibility for learning," it can be discerned that they indeed envisaged feedback to have a cyclical character since only such feedback can generate additional approaches to further improve on the same task. The purpose of feedback (from the perspective of students) also represents a central aspect of this study which was addressed primarily through the sixth and potentially also seventh research questions (see Section 2.10), and explored through a number of questions in the questionnaire (see Appendix 3) and interview schedule (see Appendix 4).

Hattie & Timperley, furthermore, distinguished four primary levels or types of feedback: feedback about the task, feedback about the processing of the task, feedback about self-

regulation, and feedback about the self as a person, each of which may answer the questions above with varying degrees of effectiveness. Feedback about the task – which may be simple or complex, related to one or several individuals and take the form of alphabetic or numeric notation – is essentially corrective information that allows the student to understand how well they are accomplishing a task. Examples of this feedback type include details on correct and incorrect answers, obtaining more or different information, or increasing knowledge on a more superficial level.

Hattie & Timperley pointed out that such feedback can be powerful and that "Having correct information is a pedestal on which the processing and self-regulation is effectively built" (p. 91). However, they warned against providing feedback about the task in too minute detail, as this may lead to a decline in students' performance in terms of cognitive effort and long-term learning strategies. They also cautioned that this type of feedback is often not generalisable to other tasks or questions, which may limit its overall value.

In contrast to feedback about the task, feedback about the processing of the task relates to deeper understanding which "involves the construction of meaning (understanding) and relates more to the relationships, cognitive processes, and transference to other more difficult or untried tasks" (p. 93). It, thus, stands on a higher plane of learning than feedback about the task since it allows for its recipients to develop the metacognitive skills that are especially necessary for self-assessment and self-regulation; it is strongly formative. Hattie & Timperley affirmed, however, that these two types of feedback can reinforce each other and have an all-the-more powerful effect on their recipients: a heightening of task confidence and self-confidence brought around by feedback about the task may result in improved information and strategy-searching skills.

Feedback about self-regulation stands above the two previous types of feedback on the uppermost level of cognition since it involves the systematic provision of corrective information for the self by the self. “It implies autonomy, self-control, self-direction, and self-discipline” (*ibid*); it is, thus, essentially self-assessment. Hattie & Timperley described two chief aspects of self-assessment: self-appraisal and self-management. Inherent in the former element is the evaluation of one’s skills, knowledge and thought processes while the latter element is associated with monitoring and regulating one’s learning behaviour.

At this level of pedagogical self-regulation, the students themselves decide on which level to assess themselves (i.e. criterion-, norm- or self-referenced) and whether they wish to seek external feedback. Hattie & Timperley also pointed out how pivotal the type of feedback provided is in having a constructive or destructive effect on how students perceive their performance and on their feelings of self-efficacy. Accordingly, students need to be able to relate it to the cause of an error or their poor performance rather than to the error or their poor performance itself. Similarly, success feedback needs to be directed specifically to what was done well so that the success can be repeated in the future.

Hattie & Timperley branded feedback about the self as a person as the least beneficial. It often takes the form of personal evaluations – usually positive, but sometimes also negative – that appear to make broad, simplistic evaluations about the student themselves. Phrases such as “You’re a great writer” or “Poor effort” are exemplar. Such feedback is too vague and lacks the informative value needed to identify those aspects of a task which were completed well or those which require further attention, or to change learning processes. The authors argued that only when such dichotomous

feedback is combined with details related to effort, self-regulation, engagement or other specific processes can it have a positive effect on self-efficacy and, eventually, on the task itself.

An item in the interview schedule asked whether the feedback students believed to receive most frequently referred to the task, to themselves or to their peers. The outcome is discussed in Section 4.3.2. This was done in an effort not only to overlap with Hattie & Timperley's feedback about the task and feedback about the self as a person, but also to gauge whether the feedback the participants received was perceived as being most frequently criterion-, norm- or self-referenced. As explained in Section 3.9, the authors' other two feedback levels were not explored for language-related reasons.

Hattie & Timperley also briefly discussed a small number of feedback issues, including the timing of feedback, the effects of positive and negative feedback, certain aspects of applying feedback in classroom situations, and using assessments for feedback. Their primary point was that there are differential effects on learners according to the level on which feedback is given. Due to the added complexity and workload for the participants, feedback was not investigated explicitly or comprehensively on the four levels the writers identified in the present study, but rather more generally.

The first two of the authors' additional feedback issues were explored, however. The timing of feedback, which was addressed by the fourth and fifth research questions, was investigated through a number of items in both the questionnaire and interview schedule. The results are discussed in Section 4.3.4. The effects of positive and negative feedback, which were seen as helpful in answering the seventh research question, were investigated through an item in the interview schedule. The results are discussed in

Section 4.3.3. In addition, the overall effect of feedback on three affective aspects in addition to one cognitive aspect, which are located within the literature in Section 2.6, was examined through an item in both the questionnaire and interview schedule. These results are discussed in the same section.

Despite the findings of the authors' works reviewed above, it appears that the relationship between feedback and its cognitive and affective impact on learners remains enigmatic or highly unpredictable at best. In fact, Koestner, Zuckerman & Koestner (1987) demonstrated in a study with undergraduate students that, while task-involvement increases intrinsic motivation more than ego-involvement, the context within which the feedback was given produced differential effects. They found that effort-focused feedback in the task-involved (game-like) condition and ability-focused feedback in the ego-involved (test-like condition) led to increased intrinsic motivation. They, thus, concluded that "the form in which praise is delivered, as well as the context in which it is received, both affect its motivational consequences" (p. 389).

Hattie & Timperley (2007), too, argued that the cultural context, or "climate," should be a serious consideration when examining and attempting to understand how feedback is given and, in particular, how it is received by learners. What may be classed as constructive feedback in one setting may be deleterious in another. Wiliam (2011) also argued that "feedback cannot be evaluated without also taking into account the instructional context in which it is provided, and used" (p. 12). For instance, in China, where a collectivist mentality dominates (McLeay & Wesson, 2014), critical comments, however well-intentioned, may, according to Carson & Nelson (1996) and Phuong-Mai, Terlouw & Pilot (2005), produce feelings of humiliation and jeopardise group harmony.

This disclaimer also needs to be made on the individual level. Carless (2006), for example, noted that "One person's helpful suggestion may be another person's judgemental criticism" (p. 229). Similarly, Wiliam (2012) cautioned that "Feedback given by a teacher to one student might motivate that student to strive harder to reach a goal, whereas exactly the same feedback given by the same teacher to another student might cause the student to give up" (p. 33). The author explained that "Even the relationship between the recipient and the person giving the feedback" (pp. 32-33) may influence the way in which learners will react to feedback. Therefore, in order to better understand how feedback impacts on learners, other contextual variables need to be taken into account.

Due to the attention in the literature given to the four aspects reviewed above, it was decided to include them in the present study in terms of how students perceived them to be influenced by the feedback they received. However, it is important to note that, with the exception of achievement, some changes were made in the way these aspects were presented conceptually to the participants. As pointed out by Harlen & Deakin Crick (2002), motivation is a multifaceted concept. It was decided to investigate it more broadly as it is defined by *Cambridge Dictionaries Online* and *Merriam-Webster Online Dictionary* in order to alleviate the linguistic and, more generally, the cognitive burden on the respondents. After piloting the questionnaire, the phrase "self-efficacy" was replaced by "self-confidence" since the former led to significant confusion amongst the pilot participants. However, the actual phenomenon investigated remained true to Bandura's (1997) definition of self-efficacy above, given the way it was explained in the questionnaire. Due to the phenomenological complexity of self-regulated learning as well as the difficulty of explicating its meaning in English to non-native English speakers, it

was decided to investigate students' studying habits as a related phenomenon and a component of self-regulated learning (Butler & Winne, 1995).

2.7 Feedback in Higher Education

It is appropriate to explore feedback specifically within higher education since this represents the study's sectoral focus. A number of high-ranking journals are dedicated to or have as one of their core foci assessment in this sector, including, for example, Assessment & Evaluation in Higher Education, Assessment in Education: Principles Policy and Practice, Assessment Update, and Educational Assessment, Evaluation and Accountability. McLeay & Wesson (2014) confirmed that assessment practices, and feedback more specifically, are gaining increasing attention amongst academics and professionals due to stiffening competition within higher education. More specifically, this includes better "understand[ing] students' expectations, perceptions and experiences" (p. 1) which represents the same key undertaking of this study. Since the literature on feedback within higher education is extensive, this section will concentrate on some of the key issues of contention and gaps in the literature identified by a number of leading scholars. This provided a useful means by which to consider how the present study could make a contribution to knowledge.

Nicol, Thomson & Breslin (2014) described feedback in higher education as a "troublesome issue," with national surveys in Australia and the U.K. showing students as being less satisfied with feedback than with any other facet of their courses (p. 102). According to Nicol & Macfarlane-Dick (2006), one common problem is that "most students have little opportunity to use directly the feedback they receive to close the performance gap" (p. 213). Indeed, even if the feedback is timely, specific and individualised – as, for example, Kluger & DeNisi (1996), Gibbs & Simpson (2004-2005),

Shute (2008) and Boud & Molloy (2013) argued it should be –, it would be of little use if students did not have the opportunity to act on it. As pointed out through definitions of feedback from Ramaprasad (1983) and Sadler (1989) in the previous section, feedback can only be denoted as such if it is used to alter the gap between the actual and desired performance. Otherwise, that information is, in the words of Sadler, but "dangling data" (p. 121).

Mirroring Nicol & Macfarlane-Dick's (2006) finding, Jonsson (2012), in a review of 103 articles related to students' use of feedback in higher education, reported that "there is ample evidence of both anecdotal and scientific nature that a number of students do not use the feedback they receive, and therefore do not realize the potential of feedback for learning" (p. 64). He suspected that this may be due to a dissatisfaction with the quality of the feedback. Indeed, the discontent of university students – especially or at least in Australia, England and Hong Kong – with the feedback they receive has been growing in recent years (Hounsell, 2007; Hounsell *et al*., 2008; Carless *et al*., 2011). They frequently lament that it is "too brief, too negative, too difficult to decipher or to understand" (Burke, 2009, p. 42). Jonsson (2012) maintained, however, that the likeliest explanation is that students do not know how to use the feedback they are given. He also characterised studies investigating students' strategies of using feedback as well as their perceptions of feedback as underdeveloped.

Evans (2013) produced a meta-analysis of 460 articles published between 2000 and 2012 by authors from Australasia, the U.K., North America, and a small number of countries in mainland Europe and Southeast Asia on the subjects of assessment and feedback in higher education. It highlighted some of the gaps and other issues specific to feedback in higher education. Due to its high relevance within this section as well as

its recent publication date, it will be afforded greater attention here. Evans paraphrased Ferguson (2011) as saying that feedback

> is seen as a crucial way to facilitate students' development as independent learners who are able to monitor, evaluate, and regulate their own learning, allowing them to feed-up and beyond graduation into professional practice (p. 72).

Therefore, feedback in higher education is important, if not, critical since it can bring about the self-regulatory behaviour which itself is an invaluable transferrable skill. It can also be gleaned from the above quotation that feedback should be employed to give students a greater sense of empowerment. Nicol & Macfarlane-Dick (2006) wrote on this point, affirming that, "in higher education, formative assessment and feedback should be used to empower students as self-regulated learners" (p. 199). In his discussion on "sustainable feedback" and speaking on the point of "the realities of mass higher education," Carless (2013) also argued for students' reduced reliance on tutor feedback.

According to Nicol & Macfarlane-Dick (2006), the reality in higher education, however, is that "formative assessment and feedback are still largely controlled by and seen as the responsibility of teachers" (p. 200). That notwithstanding, self-regulation is vital, and assessment and feedback are key agents in achieving it. The effects of feedback on students' self-regulated learning were not investigated in this study since it was felt that the vast majority of the participants would not have the requisite English proficiency to express themselves adequately on such a complex topic.

Evans (2013) also observed conflicting accounts of the improvement of feedback practices in higher education as well as a lack of substantive evidence as to the type of feedback that works best in that sector. She identified dissatisfaction with feedback from

both students and lecturers. Students complain about the technical aspects of feedback, including content, the organisation of assessment activities, timing and the lack of clarity about requirements while lecturers lament students' lack of engagement with and capitalisation on the feedback they are given.

Evans further reported a lack of studies focusing on feedback from the perspective of lecturers and postgraduate students, and an inadequate number of "true experimental designs" (p. 76). The majority are smaller-scale ($n < 100$), and also monothematic, short-term and opportunistic. Those that include interventions are inadequately detailed, lack measurability in terms of their effectiveness outcomes, and are too methodologically heterogeneous to practically allow for a systematic testing of approaches across different contexts. In addition, Evans found that a number of areas were severely underrepresented in the articles she included: only 4% investigated individual needs, 4% performance, 3% the feedback gap and 3% the affective domain.

Evans also reported disagreement on the most suitable mode, volume and timing of feedback. Moreover, while some writers have focused on how lecturers should feed back to their students to better support their learning needs, others have stressed the importance of aiding students in developing the necessary skills to become more effective self-assessors and self-regulators. Questions regarding the nature and appropriateness of such support offered to students have also been raised.

How, then, can the present study make a contribution specifically to knowledge on feedback within higher education in light of the gaps and other issues identified by the writers cited in this section? First and foremost, it aimed to shed some light on the feedback practices in higher education in the Chinese context from the students'

perspective and on which types of feedback these Chinese university students preferred. As pointed out at the outset of this chapter, literature in English on this topic appears to be sparse.

Further practical and more specific contributions this study was intended to make include, for example, the inclusion of at least a small number of postgraduates in the sample. In addition, according to Evans' numerical delineation of the scales of studies, the study – with over 200 questionnaire respondents – was not small-scale. Moreover, it investigated, *inter alia*, some of the psychological impacts of feedback, thereby contributing to one of the most underrepresented areas: the affective domain.

2.8 Formative Feedback and Feedback for Learning

As this literature review nears the study's conceptual and geographical foci, it is necessary to first examine a particular variety of feedback: formative feedback. A first glimpse of formative feedback can be caught in a work by Wiener (1950), regarded as the founder of cybernetics and a key contributor to the formalisation of the notion of feedback, in which he wrote:

> Feedback is the control of a system by *reinserting* into the system the results of its performance. If these results are merely used as numerical data for criticism of the system and its regulation, we have the simple feedback of the control engineer. If, however, the information which proceeds backwards from the performance is able to *change the general method and pattern of the performance*, we have a process which may very well be called learning (emphasis added, p. 71).

Wiener used the word "reinserting" to imply that the results (or information) of a particular system (or process) are recycled and inputted once again into that system. In doing so, this information, rather than only reflecting on the system, may actually modify

it in some way. It is this potential for modification of an existing process that lends the feedback a formative character. This understanding of feedback is reminiscent and possibly an early foreshadowing of the notion of feedback loops espoused, in particular, by Ramaprasad (1983) and Sadler (1989).

Typically, the earliest reference made to feedback within the humanities in most of the literature is Kulhavy (1977) who spoke of feedback as a continuum which may range from a simple, binary response, such as "Yes-No," to very complex content which may include significant corrective and even supplementary information. He noted that, as the complexity of the feedback increases, "the process itself takes on the form of *new instruction*, rather than informing the student solely about correctness" (emphasis added, p. 212). Although instruction is not synonymous with learning, feedback does imply the presence of an agent and a principal, usually students and teachers (Sadler, 1989). In addition and more to the point, "new instruction" signifies an iterative process whereby improvements are made to, for example, teaching methods or the content of a course. Hence, feedback, thus conceived, is formative.

Black & Wiliam (1998a) made frequent mention of formative feedback. However, they intentionally used this concept in tandem with the term "formative assessment" since the latter was not yet being widely used in the literature. Torrance (2012), who drew attention to the definitional variability of formative assessment, also used the term "formative feedback" on several occasions, at times and confusingly in the same breath as the term "formative assessment." However, it appears he considered formative assessment as a process and formative feedback as a product or as one part thereof by writing: "*engaging* in formative assessment, and *providing* formative feedback […]" (emphasis added, p. 327).

Fluckiger *et al.* (2010) argued that, in order for formative feedback to achieve maximum positive effect, it should be given *vis-à-vis* several aspects of learning, including on the product, process and progress thereof. They also asserted that, in addition to being specific, simple, descriptive, focused on the task and based on clear criteria, effective formative feedback should originate not only from the instructor but also from the students themselves as self- and peer-assessment. They maintained that "Formative feedback involving students as partners is a key strategy to enhance the teaching and learning process" (p. 140).

An especially explicit and in-depth discussion on formative feedback can be found in a frequently-cited review of literature on feedback by Shute (2008). She defined formative feedback as "information communicated to the learner that is intended to modify the learner's thinking or behavior for the purpose of improving learning" (p. 154) where "improving learning" denotes an "increase [in] student knowledge, skills, and understanding in some content area or general skill" (p. 156).

Thus, like formative assessment, the information provided to the learner through this process is intended not as a summative judgement, but specifically for purposes of educational progression. The words "thinking" and "behaviour" demonstrate the centrality of the cognitive or psychological domain in determining whether the feedback given results in long-lasting improvement in learning. Shute added that formative feedback could also apply to teachers for purposes of improving instruction but that the primary recipients of formative feedback were learners.

Shute likened formative feedback to scaffolding which initially supports learners in attaining higher-level thinking and problem-solving skills before being gradually removed,

leaving learners with greater autonomy to perform functions previously carried out solely by teachers. She outlined three primary uses of formative feedback for learners. First, it can provide information to learners about the gap between their current level of performance and that which is desired, which can ease uncertainty and increase motivation. Second, it can be especially effective for lower-achieving learners and be used as a support device to reduce their "cognitive load," as she termed it. Finally, it can be used to correct ineffective learning methods, particularly when it is more specific.

Shute compared effective and useful formative feedback to "a good murder" in that it requires a motive (a reason why the student would need it), opportunity (timeliness for it be useful for the student) and means (ability and motivation on the part of the student to use it) (p. 175). She cautioned, however, that, even given these three elements, there is significant variability in terms of the effects of feedback on performance and learning. She also stressed again that "there is no 'best' type of formative feedback for all learners and learning outcomes" (p. 182), but endorsed formative feedback generally in terms of improvement in learning as well as teaching.

Despite her admission that there is no optimal type of formative feedback, Shute stressed that "feedback that has negative effects on learning is not formative" (p. 156). Accordingly, formative feedback is and can only be associated with positive effects. However, following the understanding of the word "formative" argued in Section 2.5, this concept cannot necessarily be associated with positive effects. Thus, it is reasoned that the phrase "for learning" rather than "formative" is a more accurate embodiment of the objective of formative assessment as advocated throughout the literature.

It appears that feedback for learning has not (yet) established itself as a stand-alone concept in the literature. A search produced only a very small number of instances of its usage, primarily in websites. The few publications which were found using this concept – including Askew (2002), Hendry, Bromberger & Armstrong (2011) and Jonsson (2012) – appear to use it not as a type of feedback in its own right but rather more broadly to refer to that which supports learning. The only deliberate or systematic usage that could be found is by Boud & Molloy (2013) who, however, do not appear to use it as a separate concept *per se* but, instead, to disassociate it from feedback. In their view, feedback for learning repositions feedback

> (1) From an act of teachers to an act of students in which teachers are part [*sic*] (from unilateral to co-constructed; from monologue to dialogue); (2) From the almost exclusive use of teachers to that of many others (from single source to multiple sources); (3) From an act of students as individuals to one that necessarily implicates peers (from individualistic to collectivist); [and] (4) From a collection of isolated acts to a designed sequence of development over time (from unitary items to curriculum) (p. 710).

Though this conceptualisation may be useful in terms of dissecting feedback, it contrasts with the understanding of feedback for learning proposed above. Using words like "act" and "use," it focuses heavily on the process rather than on the outcome. Furthermore, it bears a close resemblance to Stiggins' (2002) understanding of Assessment for Learning in which he underscored that students should be as much a part of the assessment process as teachers, and that there should be a continuous flow of information about their learning.

Despite feedback for learning not being a prevalent concept in the extant literature, and although it does not appear explicitly as one of this study's research foci, it does more justice to the phenomenon being investigated than formative feedback. This is in

consideration of how the word “formative” has been conceptualised in this thesis. The specific nature of the type of feedback explored in the study is illustrated, for instance, by the words “helpful,” “prefer” and “effective” contained in the research questions which point to the participants’ experiences of and expectations towards feedback that they perceive not only to have an altering but also a positive impact on their learning.

A further motivation for this conceptual choice was a publication from Stiggins (2005) in which he argued for a differentiation between formative assessment and Assessment for Learning. He asserted that, “When used effectively, assessment FOR learning *always* triggers an *optimistic* response to assessment results from within the learner” (capitalisation in original, emphasis added, p. 328). Thus, according to Stiggins, “for learning” is equivalent to positive outcomes in terms of the effects of the feedback or assessment activity on learners.

The same understanding will be used here. However, since the present study focused on feedback, not the term “assessment” but the concept “feedback” (as one element within a broader system of or approach to assessment) was paired with the phrase “for learning” to represent the broader phenomenon under investigation. Feedback for learning, however, is not intended to represent a wholly new form of feedback. It embodies the same central tenets as formative assessment and formative feedback, but, as alluded to above, more faithfully reflects positive outcomes of modifications to learning and teaching processes. For purposes of data collection, only the generic term “feedback” was used, and was defined only as marks or grades, oral or written comments (see Appendix 1) for ease of understanding on the part of the participants.

2.9 Formative Assessment & Feedback in China

This literature review has arrived at the study's geographical and cultural focus: China. Before engaging with individual studies related to formative assessment and feedback in China, it may be useful to gain an impression of the standing formative assessment enjoys in China at governmental level. A number of steps taken by the Chinese government to develop and implement formative assessment in its public school system are also summarised.

2.9.1 The Official Position

The concept of formative assessment, and proposals to implement it in China's educational system, first emerged nearly three decades ago. It was then that the so-called "Syllabus of Moral Assessment" for primary and secondary schools was published which delineated five guidelines aimed at both instruction and assessment. According to one of these guidelines, assessment should be "conducted democratically, combining self-assessment, peer/group appraisal with teachers' comments" (SEC, 1988, quoted in Wang, 1996, p. 83). Considering the role of the teacher as the key figure in all matters in Chinese classrooms that prevails even today, this statement is momentous, as it points to a sharing of assessment activities.

The year 2001 saw a major development in Chinese education history. It was then that a document detailing measures for the reform of China's basic education system was published by the Chinese Ministry of Education. The majority of the reforms centred around tackling illiteracy, restructuring and consolidating the higher education sector, and increasing funding for education (Li, 2004). Adjustments to the curriculum were also proposed, including in the area of assessment. One of the aims of the Chinese Central Government at that time was to allow "assessment […] to stimulate learning. The new

curriculum," as it was envisaged, "encourages self-, peer-, and parental assessment in addition to teachers'" (Chang, 2002, quoted in Berry, 2011, p. 52).

In another guideline document issued by the Chinese Ministry of Education in 2007 entitled *College English Curriculum Requirements*, it is stated that "assessment of College English learning should include both formative assessment and summative assessment" (cited in Chen *et al*., 2012, p. 1). In the same document, formative assessment was defined as:

> the procedural and developmental assessment conducted during the process of teaching and learning … It is a means to adapt various assessment approaches and means to follow up the teaching and learning process, and to provide timely feedback information so as to enhance students' overall development … Formative assessment includes self-assessment, peer-assessment, teachers' and the administration's assessment of students' learning … It is used to observe, evaluate and monitor the learning process for the purpose of enhancing effective learning (cited in *ibid*, pp. 6-7).

From this explication, it is clear that much emphasis is placed on timing: it should be carried out "during the process of teaching and learning" and it should be "timely." Moreover, it, again, underscores the participation of students in addition to teachers and the administration. The formative elements are manifested in the words "enhance students' overall development" and "for the purpose of enhancing effective learning." As Chen *et al.* (2012) pointed out, this understanding of formative assessment echoes that generally supported in Western contexts.

Gao (2002) reported that, in the case of secondary English teaching in China, forms of assessment have become more multifaceted. In addition:

> Formative assessment is carried out during the developmental stages of the teaching program, for the purposes of monitoring the learning process, motivating the learner, revising the materials, altering the teaching program, reconsidering the goal and objective, etc. (quoted in Wang & Lê, 2006, p. 4).

The English Language Curriculum Guidelines for Senior Secondary School [*sic*], published in 2003, signifies another step towards the formalisation of formative assessment within the Chinese education system. The document recommended that "both formative assessment and summative assessment be used and attention be paid to the motivating and facilitating roles of assessment in students' learning" (Ministry of Education, 2003, quoted in Wang & Lam, 2009, p. 74). However, these guidelines were limited to English and to the secondary level. Another initiative from the Chinese Ministry of Education in 2007 launched the "College English Reform Program" which aimed at reforming the assessment framework by promoting the incorporation of formative assessment alongside summative assessment (Chen & Klenowski, 2009). However, as the name implies, this was also limited to English and to the tertiary level (i.e. colleges and universities).

Developments in assessment are, however, not restricted to English learning and teaching. A notification document released by the Chinese Central Government entitled "On the implementation of assessment reforms of primary and secondary schools," for example, required schools "to use AfL as a major focus in their […] educational planning" (Berry, 2011, p. 52). It also called for the assessment methods teachers employ to be more varied and for them to include "open-ended assessment items" in order to acquire a better appreciation of students' needs, their ability and of how to support them on a more individual basis (p. 53).

According to Berry, the most promising indications of formative assessment being put into practice could be seen in the southern and most populous Chinese province of Guangdong. There, a project foresaw the combination of formative and summative assessment. On a trial basis, authorities in the provincial capital of Guangzhou required judgements of the performance of new senior secondary students to be made both summatively and formatively. The author, however, did not provide further details on how this could be achieved.

By looking at the above reports, it does appear that the Chinese Central Government has both acknowledged the importance of formative assessment and made overtures to implement it within the national education system. The questions remain, however, whether any of these policies have been translated into practice at the classroom level and, if so, whether they, as Marshall & Drummond (2006) put it, embody the true "spirit" of Assessment for Learning or merely conform to the "letter" (p. 137).

The evidence points to this not being the case. Wang (2008), for example, although writing of formative assessment as being "established" in secondary schools in China, conceded that "the reform of education or curriculum evaluation is in form rather than in real meaning" (p. 234). She corroborated this view with 100 questionnaire responses from junior middle school English teachers in the north-eastern province of Shanxi. Some of the respondents reported practical and logistical limitations, including large class sizes and inadequate time for professional educational assessment development due to restrictions brought about by requirements for high pass rates. Others criticised the complexity of formative assessment which they felt lowered their interest and incentive to use it in the classroom.

Wang's findings are reinforced by a survey conducted by Jin (2010) which investigated, *inter alia*, the content of language testing and assessment courses of 86 different universities countrywide. She determined that assessment-related topics "did not have sufficient coverage" (p. 567). Echoing Wang (2008), Zhan & Wan (2010) observed that the Chinese Ministry of Education has, in recent years, supported Assessment for Learning, notably in English language education, but admitted that tangible reform of China's assessment culture, which continues to be exam-oriented, is still lacking. Chen *et al*. (2014) blamed the "top-down nature" of government-led initiatives, inadequate sensitivity to local circumstances and cultural differences as well as insufficient teacher training and professional development (p. 282).

Jin (2008) lamented that "most college English teachers in Mainland China only pay lip service to formative assessment and consider that its practice is unrealistic in the present educational context" (quoted in Zhan & Wan, 2010, p. 9). It goes without saying that such a statement – "most English teachers in Mainland China" – cannot be accurate, as that would represent an enormous number of individuals. Despite this, particularly the words "unrealistic in the present education context" may point to much more fundamental and deep-rooted barriers to the implementation of formative assessment in China. Whether this applies only to English or to all subjects, or whether the situation is different at the primary or secondary levels is uncertain. The majority of literature in English that could be found on formative assessment in China deals specifically with English learning and teaching, and with that at the tertiary level.

One possible (other) source of the problem may be definitional. In their study, which compared and analysed the assessment practices of university English as a Foreign/Second Language instructors in Canada, (mainland) China and Hong Kong,

Cheng, Rogers & Hu (2004) defined assessment as "the process of collecting information about a student to *aid in* decision-making about the *progress* and language *development* of the student" whereas they understood evaluation as "the interpretation of assessment results that describes the *worth or merit* of a student's performance in relation to a set of learner expectations or standards of performance" (emphasis added, p. 363). Thus, it follows from their conception that assessment is of a more formative nature while evaluation carries a more summative connotation.

However, in both their discussion and their data collection tools examining teachers' views on the purposes of assessment and evaluation, the authors habitually combined these two concepts. For example, in Table 1 entitled "Purpose of assessment and evaluation," two of the student-centred purposes they included are to "Provide feedback to my students as they progress through the course" and to "Determine final grades for my students" (p. 367). As shown in earlier discussions in this chapter, feedback is typically associated with formative purposes whereas determining final grades is, by definition, a summative exercise.

Another example can be seen in the next category, instruction-centred purposes, wherein one purpose is to "Diagnose strengths and weaknesses in my own teaching and instruction" and another to "Group my students at the right level of instruction in my class" (*ibid*). As with the first example, the former purpose, particularly owing to the word "diagnose," lends itself to formative functions; the second one, however, is clearly more summative in nature since classifying students into specific levels is a process of selection and akin to giving them a rank or grade.

In their conclusion, Cheng, Rogers & Hu noted that, “As with any mail surveys, there is the potential that not all instructors interpreted the survey questions in the same way” (p. 380). This will indeed likely have been the case, not only because their survey was distributed through the post, but also because two key terms with different meanings were used interchangeably in the questionnaire (see their Appendix: pp. 383-389).

This may point to a misunderstanding or to a different interpretation of assessment concepts which could be shared by several Chinese scholars and educationists. A study conducted by Shing & King Fai (2007), which looked at Chinese college lecturers’ conceptions of assessment, offers some insights into the problem of terminological interpretation. Their overall conclusion was that the college lecturers they surveyed held “very contradictory conceptions of assessment:” while they agreed that assessment improves teaching quality, student learning and school accountability, they held that it could not validly describe student learning and that it was inaccurate (pp. 194-195).

In an effort to explain this incongruity, the authors differentiated between a “functional” and an “evaluative” understanding of assessment. They stated that the former term “refers to the situation in which assessment urges teachers to train their students better in order to enhance performance in examinations” while the later term “refers to the analysis of how well the assessment results reveal what they intend to measure” (p. 196). The writers, thus, appear to align the concept of evaluation with test validity. Furthermore, evaluation, rather than being horizontally juxtaposed with assessment, is what comes *after* assessment. In other words, according to the authors, an assessment is evaluated, and it is evaluated for its validity.

Though contradicting the earlier preliminary conclusion that the Chinese Central Government has made efforts to put formative assessment into practice, the problem may, alternatively or additionally, be one of lacking governmental or municipal enforcement. Supporting this, Wang (2008) stated that "The Ministry of Education in China has not established effective evaluation mechanisms which support curriculum reform and development" (p. 4). Another extract from Dello-Iacovo (2009) casts further doubt over the curriculum reforms having taken root or being heeded by educational institutions in China:

> examination content remains in line with traditional rote learning methods and has significantly undermined the reforms. […] In Beijing the municipal and county level examinations were all still written in the style, focusing on the memorisation of knowledge, even including content not in the textbooks. The examinations were often completely at odds with the prescribed curriculum. For example primary English lessons were supposed to focus on listening and speaking but were examined with a written test which diverged greatly from the curriculum requirements and Ethics examinations focused on memorisation rather than student values (p. 247).

Although the above extract may carry considerable currency in terms of revealing assessment practices in 21st century China owing to the fact that it reports findings derived from the country's capital, Beijing, any account or study on China should be treated with caution. China is a demographically as well as geographically vast country; even studies drawing data from hundreds of schools and thousands of individuals would very likely not be representative of the rest of the theoretical population. There are over 31 million students at the tertiary level alone (United Nations Educational, Scientific and Cultural Organisation Institute of Statistics, 2012), attending over two and a half thousand colleges and universities across the country (Ministry of Education of the People's Republic of China, 2014).

Even two decades ago, Manion (1994) contended that, “for most researchers, obtaining a nation-wide probability sample of the Chinese population is both impossible and impractical” (p. 741), justifying this not only on the grounds that China has an enormous population and a large number of different cultural groups, but also because of the limitations which exist on particularly foreign researchers brought about by a more closed political system. Therefore, diverging accounts of particular practices within China’s education system are to be expected and prudence should be exercised when considering the representativeness of conclusions drawn from them.

With this in mind, it is difficult to ascertain whether formative assessment has yet been widely implemented in Chinese classrooms. This would not be expected since some authors have argued that this is still not the case even in the birthplace of the Assessment for Learning movement, the U.K. (Pryor & Crossourard, 2008; Mansell, James & Assessment Reform Group, 2009; Boyle & Charles, 2010; Hill, 2011) – a country far smaller than China, and with a vastly different history and educational culture. When the U.K. government’s Assessment for Learning Strategy was already in its third year, Black himself admitted that Assessment for Learning "’isn't happening’ in a ‘very large number of classrooms’" (quoted in Stewart, 2010, para. 3).

Another initiative of the U.K. government – Assessing Pupils’ Progress – has also been criticised. Developed by the Qualifications and Curriculum Development Agency together with National Strategies, it aimed at enabling primary and secondary schools in England and Wales to apply Assessment for Learning. Swaffield (2009) described it as a “(mis)interpretation of AfL as a teacher driven mechanism for advancing students up a prescribed ladder of subject attainment” (p. 6), that is, as being linked with summative rather than formative assessment.

Similarly, Stobart (2010) commented that the Initiative "claims to be AfL put into practice, yet is more about standardising teachers' summative assessment than about day to day classroom learning" (p. 2). He exemplified this by writing that "publishers are marketing materials and tests branded as AfL which are little more than repackaged summative test materials" (*ibid*). He wrote of Assessment for Learning as, thus, experiencing an "identity crisis" whereby it was at risk of losing its focus and, whereas once it was "trim," it had become "flabby" (*ibid*).

Thus, some government programmes which are believed or touted as being successful may, at least from academics' point of view, not necessarily be so. This likely also applies to other countries, including China. Considering the large number of local educational authorities, schools, teachers and individual students in China, it is likely that snapshots of the Chinese education system will differ significantly.

It may, therefore, be more fruitful to focus on individual studies to see how formative assessment and feedback were received in those settings in China in which they were specifically investigated or implemented. This will help shed some light on the question of their compatibility with a range of Chinese contexts. In order to maintain closer relevance with the present study, however, only those studies that explored the effects of formative assessment and feedback on students (from all levels of education) or their views thereto have been included.

2.9.2 The Academic Perspective

A study carried out by Wei & Chen (2004) looked at the implementation of self-assessment by 90 mostly male Chinese learners of English in a writing course at a prestigious university in Shanghai over a one-year period. In order to assess themselves,

students were given two checklists: one with error symbols and another which included writing strategy questions on content, organisation and language. Students were also assessed by their teachers and peers. Through their observations and analysis of the comments produced through the assessment exercises, the researchers concluded that the self-assessment technique they had developed and trialled led to a reduction in teacher workload and an improvement of learning efficiency. They reported that, although attitudes were generally positive, students felt that some writing skills were easier to improve than others and that it was not possible to improve to the desired level through self-assessment in the case of English writing due, primarily, to limited language competency.

Another study carried out by Wang & Lê (2006) involved a total of 540 students at a middle school in the north-eastern province of Shanxi where oral formative assessment was being implemented within the English language classes through a number of classroom activities, including Question and Answer, reading aloud, role plays, singing songs, dialogues, games, interviewing and reporting. Teachers observed and monitored students' learning behaviour and provided feedback both to them and their parents. Although the authors discussed a variety of formative assessment techniques, such as self-assessment, teacher assessment, peer assessment, teacher and student joint assessment and parent-based assessment, it is not clear whether all or any of these techniques were actually employed during the study.

The researchers' analysis of the findings focuses on the teachers' reactions rather than on those of the students or their parents. In addition, only a small number of excerpts from the interviews with the teachers were included and no explanation as to the way in which the coding was done was provided. The above notwithstanding, assuming

accurate interpretation of the teachers' comments as well as of the teachers' interpretations of the students' reactions to the utilisation of formative assessment, the students' interest, learning motivation, attainment and self-efficacy were all heightened. Teachers, however, also admitted to a lack of understanding or a misunderstanding of the role and applications of assessment in general. Formative assessment was also deemed by some as unsuitable in an assessment culture which, as noted in Sections 1.3 and 2.9.1, is still heavily examination-orientated.

Qu & Yang (2010) implemented self- and peer-assessment activities through group presentations, involving 33 graduate students at a university in Beijing. The participants were provided with an assessment rubric, and were tasked with rating other groups' presentations and providing feedback which was later shared with all the groups. The teachers also gave feedback to individual students who performed poorly and to the class as a whole. In addition, students were asked to create a written composition in which they rated and reflected on their own as well as their group's performance. Thereupon, the students shared their reactions to the self- and peer-assessment activities with their teachers through group interviews.

The results were positive. The students reported that, in addition to heightening their interest, assessing their peers caused them to take the task more seriously. Qu & Yang quoted one student as saying: "Rating is like a mirror which allows us to see our own mistakes and know which mistakes could be avoided" (p. 778). This comment illustrates that the assessment activity not only activated self-reflection but that it may also have corrected a potential flaw or deficiency in that student's *future* learning processes. Hence, this particular assessment activity may be considered as strongly formative.

An action research study conducted by Wei (2010) within an English course at a university in the southern province of Guangxi, covering three semesters and involving 227 students in total, also involved self- and peer-assessment. However, the goal was to blend "FA [(formative assessment)] and summative assessment, making FA and summative assessment weigh half respectively in the final score portion of the English course" (p. 839). Data were collected through a large range of instruments: questionnaires, interviews, portfolio scripts, final reflection report scripts and English final examinations. While not quoting any of the students' comments, the author's overarching findings indicated that these assessment techniques had: strengthened their intrinsic motivation; increased their awareness of their current level and of what they needed to do or how far they needed to go to achieve a certain level; enabled greater variety in learning strategies and more creativity; fostered independent learning; enhanced team work spirit; and improved the relationship amongst them.

It is unfortunate that Wei did not go into detail on how exactly formative assessment was used in combination with summative assessment since it is not evident how this could be done. Although Wiliam (2000) and Black *et al*. (2003), for example, left room for assessment serving both formative and summative purposes under certain preconditions, the function of formative assessment itself is, as discussed in Section 2.4, essentially to improve learning and not for grading purposes, be it wholly or only partially.

A survey conducted by Wang (2010) comprised 609 participants who were randomly selected from universities in three different provinces. In addition to students and teachers, it also included deputy presidents' views. Data were gathered through a questionnaire and an interview. An aspect of this survey which gives it a special place in this literature review is the fact that it investigated feedback, specifically preferences for

particular forms of written error correction. As Ronayne (1999) and Lee (2011), for example, pointed out, written error correction and feedback, depending on how they are utilised, can be used for formative assessment purposes.

Two questions in Wang's (2010) questionnaire are of particular interest here since they address elements which are also explored by the questionnaire used in the present study. One question investigated the preferred frequency of (summative) feedback on the organisation of written compositions. Nearly all participants responded that they would either usually or always like their teachers to give feedback. However, only 18% and 12% of teachers, respectively, reported that they actually provided feedback with these frequencies. Another question queried how effective the respondents thought teachers' written error corrections were. The vast majority of the students and, interestingly, also the teachers rated it with the lowest category – "negative." Although the author did not explain why the majority of the teachers rated the effectiveness of their own written error correction so low, he did state that "Most students found it very dispiriting if their written work was covered in red ink, underlined and with words crossed out" (p. 199).

Using a questionnaire, Zhu (2012) investigated the relationship between metacognition – which she defined as "higher order thinking which involves active control over the cognitive processes engaged in learning" (p. 79) – and actual as well as preferred feedback strategies of 109 third-year undergraduate students at a university in Beijing. The study included the analysis of some of the electronic correspondence between the participants and one of their teachers in order to determine the effect of the tutor feedback on their motivation and achievement. Feedback strategies were divided into four categories: positive or negative comments not referring to specific aspects; positive

or negative comments referring to specific aspects; more constructive and elaborate comments on students' work with suggestions for improvement; and only a score or ranking.

In addition to an increase in students' motivation and achievement, the results also showed that students with low metacognition preferred more of the third type of feedback than those with high metacognition. The first and fourth types of feedback were less favoured by the participants and less frequently used by the teacher whose actual feedback strategies corresponded to students' preferences according to their metacognition level. Although the study revolved around only one teacher's feedback and while categorising students as either having high or low metacognition may be overly simplistic – limitations acknowledged in the paper –, the author deduced that the formative value of feedback can be increased by taking into consideration students' metacognition level and the adoption of suitable feedback methods.

A study from Tang (2013), conducted over a two-year period at a middle school in a major city in the southern province of Guangdong, examined 122 students' views of using a portfolio for formative purposes and improvements in their English learning outcomes through this method. In addition to completed textbook exercises, the portfolio also included a self-assessment of a number of tasks, feedback for the teacher, feedback from the parents, and a summative self-assessment. Data were gathered through a self-administered questionnaire with all participants and semi-structured interviews with a randomly selected subset of the sample.

According to the author, all those interviewed spoke favourably of using a portfolio for formative purposes, citing improvements in language, communication with the teacher

and their parents, confidence and motivation, and clarity regarding study goals. More than 80% of questionnaire respondents selected the most positive response to the questions examining whether: they liked the assessment method; it increased their confidence in learning English; it promoted communication between them, their parents and the teacher; and whether they believed it helped them make real improvements in their level of English.

2.10 Conclusions and Research Questions

The foregoing discussion provided a review of some of the major contributions on the topics of formative assessment and feedback in the literature at large as well as specifically within higher education and in China. In addition, it offered a discussion of the meaning of these two concepts, and argued for a particular understanding of the concept "formative." It also argued for the use of the term "feedback for learning." Although the study did not specifically investigate feedback for learning, that is, only positive effects of feedback, it did attempt to uncover to what extent this appeared to be the case.

A major conclusion that can be drawn from the literature reviewed within the Chinese context is that the Chinese Central Government as well as some provincial governments and local authorities appear to have recognised the importance and value of formative assessment in recent decades. However, it remains uncertain whether formative assessment has been successfully integrated within the Chinese education system as well as within the hearts and minds of Chinese students and teachers. If the reports included here are any indication and if one were to (likely wrongly) operate under the premise that conceptions of formative assessment are contextually independent, it is unlikely that formative assessment represents an integral part of many instructors'

teaching methodology in China. Despite this, when considering the studies reviewed here, attitudes towards formative assessment, particularly on the part of the students, are, generally, very positive.

With respect to specific themes emerging in the China-specific literature, self- and peer-assessment appear to be prominent foci since a large number of those articles included here examined them under experimental conditions, though the reason for this interest is unknown. However, even if all those articles included here investigated the same phenomenon, its importance within the Chinese literature and the Chinese education system in broader terms would be inconclusive given the small total number of articles. A far clearer theme, perhaps, and also more important is the lack of attention given specifically to feedback; it is the focus of only two of the English publications produced by the literature search. This provided additional motivation for feedback as the core aspect of the present study.

The overarching theme of this study – Chinese students' experiences of and attitudes towards tutor feedback – is divided into the following individual research questions:

(1) How helpful do they find each type of feedback they receive?
(2) How frequently do the participants believe they receive each type of feedback?
(3) How do they perceive the feedback they receive from their teachers to affect their academic performance, studying habits, motivation and self-confidence?
(4) At which point(s) of the assessment process do they claim to normally receive feedback?
(5) At which point(s) of the assessment process do they prefer to receive feedback?
(6) Which function(s) do they expect feedback to fulfil?
(7) How do they characterise effective feedback?

Chapter 3

Methodology and Research Instruments

3.1 Organisation of the Chapter

This chapter begins with coverage of the more theoretical aspects before transitioning to the more concrete elements of the study. The initial subsection presents and substantiates the project's positioning as a mixed-methods study and of the joint employment of the two chosen research methods. The second subsection justifies the study's focus on the qualitative element. The next subsection briefly reflects on the matter of bias and on how knowledge is produced through qualitative research. The subsequent subsection contends with the issues of reliability and validity – both generally and more specifically within qualitative research – and explains how these two concepts require reconceptualisation to better fit the realities that exist within that paradigm. The following section balances a number of methodological questions related specifically to research conducted in China and with Chinese participants.

The section opening the second portion of the chapter, dealing with the more tangible facets of the study, provides an account and brief description of the major steps involved in the study's preparation and data collection process. Information on the piloting of the research instruments is provided in the following section. The next section presents a brief characterisation of the study's sample population, and discusses and explains the chosen sampling methods. A clarification of the approach taken to the analysis of the data as well as an explanation of the strategies employed to code the data are offered in the proceeding section. Following this is a description of how the study met the relevant ethical requirements. Before a brief recapitulation of the chapter is made, a final section

provides a rationalisation and critical discussion of the chosen research tools; an overview of the questionnaire items; a more detailed description of the three open-response questions and their objectives; and a number of statistical observations related to the questionnaire data.

3.2 A Mixed-Methods Study

The study combined two methods and two paradigms. Far from being a novel idea, combining quantitative and qualitative research has been formalised to such an extent that it has become a paradigm of socio-scientific investigation in its own right (Bryman, 2006; Burke & Onwuegbuzie, 2004); there are even several different types of mixed-methods research (Creswell, 2003; Leech & Onwuegbuzie, 2009).

Delamont (2002) pointed out that two methods of inquiry can, if “genuinely combined,” also yield new insights (p. 45). Cohen, Manion & Morrison (2011) repeatedly underlined the value of this approach, stating that quantitative data alone cannot capture the complexity of human behaviour. In fact, Gillham (2007) averred that one research method is usually insufficient, especially if the researcher intends on using a questionnaire. Moreover, if a finding is triangulated through the use of multiple methods, then any conclusions drawn would likely be more meaningful. Some authors even speak of an *obligation* to triangulate through mixed methods. Mathison (1988), for example, stated that "Good research practice obligates the researcher to triangulate, that is, to use multiple methods, data sources, and researchers to enhance the validity of research findings" and that this is the case "Regardless of which philosophical, epistemological, or methodological perspectives" they are working from (p. 13).

As well as increasing confidence in the results, triangulation through the employment of

a variety of data collection tools enjoys further advantages. According to Arksey & Knight (1999), these may include: completeness as well as a finer understanding of the study; ability to address questions which are different but complementary; increased understanding of one dataset through analysis of another; and deviations between different datasets which may illuminate new facets to the investigation and lead to the positing of new theories or to the amendment of existing ones.

However, the writers also pointed out that there is a number of potential drawbacks to a mixed-methods approach, one of which is a higher investment of time and other resources. Researchers who are not as well-versed in the application of this approach may face additional challenges, for instance when carrying out duplication and comparative investigations. For the same reason, they may also feel inclined to falsely assume or assert congruence between datasets where there is none. However, with prudent allocation of resources, sufficient training and experience, and the appropriate amount of rigour, many of these challenges can be offset.

There is also the matter of validity which is a decisive factor in determining the strength of a study. Denzin (1970) cautioned that, although “the flaws of one method are often the strengths of another” (p. 308) and that, by combining them, one can have the best of both worlds, doing so will not necessarily lead to greater validity or to the triangulation of certain findings. Fielding & Fielding (1986) also called attention to the fact that “different methods have emerged as a product of different theoretical traditions, and therefore combining them can add range and depth, but not accuracy” (p. 33). Challenges to validity should still be addressed in the case of each method individually.

Although recognising the hazards inherent in combining quantitative and qualitative methods on account of their paradigmatic differences, Kvale & Brinkmann (2009) argued that this should not preclude their combination and emphasised that “The important issues of mixed methods are less on a paradigmatic level than on a practical level” (p. 117). In other words, the core problem is one of the necessary training needed for successfully working with two different methods.

What, then, speaks in favour of specifically combining a questionnaire with an interview as in the present study? There exists a natural partnership between these two methods: the former is better suited to investigating a phenomenon in terms of coverage since it can contain a large number of closed questions which can be answered quickly by respondents; the latter can better tackle the same phenomenon in terms of depth since it can contain more open-response questions which participants may answer orally and where there is greater possibility for back-and-forth communication between the researcher and participant. Questionnaires are, therefore, often combined with interviews (Gray, 2004). Arksey & Knight (1999) also argued for their complementariness since interviews can be used to substantiate or explore in greater depth potential themes encountered through questionnaires.

3.3 A Qualitative Focus

Cooper & White (2012) emphasised that "it is essential that the researcher recognizes the paradigm within which he or she is operating" (p. 2). Despite combining quantitative and qualitative research, the qualitative data captured through the present study played the chief role in deepening insight into the research topic. It, thus, also has a more prominent function in the analysis and interpretation of the findings (see Chapter 4). This section argues for the value of qualitative research and explains why qualitative

evidence is more suitable for investigating and understanding the central theme of this study.

There is an intrinsic untidiness associated with personal experiences and perceptions – phenomena which are difficult to classify and explain neatly, and which qualitative research is more apt at capturing. In a similar vein, Cooper & White commented:

> In this age of blurred genres and mixed methods, the postmodern era helps one to realize that things are not as they were. It is now more difficult to compartmentalize research into neat, mutually exclusive domains. Perhaps it was always thus and we are only now beginning to recognize that the messy, descriptive nature of qualitative research is the norm (p. 121).

According to Patton (2002), qualitative research employs a naturalistic approach whereby researchers attempt to understand phenomena in their real-world settings rather than manipulating them in laboratory environments. The word "manipulate" is key since it refers to interfering intentionally in natural processes, such as the reaction between two chemical compounds. Social processes – such as students' interactions with feedback – could also be influenced through experimental interventions. Indeed, the data drawn from them could likewise be analysed in laboratory-like settings, such as on a computer. However, such qualitative phenomena cannot be systematically manipulated to the extent that natural materials can; they involve sentient organisms with higher-order cognitive abilities and are far more complex.

The wording of the overarching research question and individual research questions (see Section 2.10) provide a first justification for this study's emphasis on qualitative data. Preissle (2006) pointed out that qualitative research "emphasizes the 'how' – procedures, techniques, practices – rather than the 'what' and the 'why'" (p. 690). Hennink, Bailey &

Hutter (2011) explained that such questions are suitable for qualitative research since they "focus on exploring processes behind behaviour to *understand* (or *Verstehen*)" that behaviour and also because they "seek to get an *insight into perceptions, opinions, beliefs and feelings*" (emphasis in original, p. 35). Agee (2009) maintained that qualitative research questions "need to articulate what a researcher wants to know about the intentions and perspectives of those involved in social interactions" (p. 432). In addition, they need to demonstrate geographical specificity or the subject-oriented nature of the study.

This study's research questions, generally, possess all these attributes. They do include interrogative words which, according to Preissle (2006), are not typical of qualitative research. However, as Creswell (2003) noted, qualitative enquiry can also take the form of "what" questions. The word "which" would also be a viable interrogative in a qualitative research question. For example, "What are their beliefs about…?" or "Which values do they hold towards…?" may also be construed as questions within a qualitatively focused study. A number of verbs, including "perceive," "prefer" and "expect," also point to the qualitative nature of the data collected.

Quantitative data, however, also played an important role. The second research question, for instance, investigated frequency which is a quantitative phenomenon. The data drawn through the fourth and fifth research questions can also be characterised as quantitative since they explored timing. These quantitative elements were included in order to enrich the data, aid in the uncovering of potential themes, and, more generally, to offer a broader perspective of the subject under examination.

Another motivation for the prominence of qualitative data within this study emanated from the role of the researcher and the interpretive aspect – elements which are mirrored in a definition of qualitative research from Denzin & Lincoln (2005): “a situated activity that locates the observer in the world” and “which consists of a set of interpretive, material practices that make the world visible” (p. 3). They went on to say that “these practices turn the world into a series of representations” and that “qualitative researchers study things in their natural settings, attempting to make sense of, or interpret, phenomena in terms of the meanings people bring to them” (*ibid*).

A further aspect of the study that merited greater emphasis on qualitative evidence was the close proximity of the researcher to the participants. Flick (2002) referred to a “diversification of life worlds” in a world characterised by “rapid social change” which confronts “social researchers with new social contexts and perspectives” (p. 2). He explained that, in contrast to quantitative research, qualitative methodologies “take the researcher’s communication with the field and its members as an explicit part of knowledge production instead of excluding it as far as possible as an intervening variable” (p. 6).

The consequence of these two features is that the context in which the study was carried out gained heightened importance. The field in which qualitative researchers work is referred to as the “natural setting” (Creswell, 2003) or, as Rossman and Rallis (2003) put it, the “natural world” (in lieu of “experimental laboratory conditions”) (pp. 8-9). Including not only the element(s) of human behaviour, but also the location itself in the investigation allows encounters with that behaviour to become more meaningful to an outsider.

3.4 On Bias and Epistemology

Judgements made on the basis of data collected from the field need to be made through a process of analysis characterised by ethical responsibility and scientific rigor (Lincoln, 2009). One of the greatest threats to these two aspects is bias which Cohen, Manion & Morrison (2011) defined as, collectively, "the subjectivity of respondents, their opinions, attitudes and perspectives" (p. 179). Bias is also sometimes referred to as subjectivity or "researcher bias" (Miles, 1990; Norris, 1997; Creswell & Miller, 2000). Ford *et al*. (2008) warned that, "From the research focus to the research questions, to the research design and methodology, to the theoretical orientation, and to data interpretation, research is riddled with subjectivity" (p. 83).

This applies, in particular, to open-response questions which allow participants to answer freely rather than being forced to select from predefined categories. As exemplified above, certain aspects of this kind of data may, of course, still be analysed in quantitative terms, such as for the frequency of certain words or phrases. However, it is likely that, given the sovereignty the respondents have over the way in which they word their responses, the language will be idiosyncratic and highly varied; one concept or idea may be expressed in a multitude of different ways.

However, even a question which investigates frequency directly may not be able to escape this problem. For example, the second item in this study's questionnaire (see Appendix 3) investigated the frequency with which the respondents received each type of feedback. The danger of answers like "once per class" or "ten times per week" being misconstrued by the researcher or the reader of their findings would have been lower than in the case of the options which were actually made available. However, since respondents could not be expected to remember precisely how many instances of each

type of feedback they experienced during any given period, broader frequency categories had to be employed. Admittedly, the response “often,” for instance, is not only subjective (i.e. dependent on the understanding of the respondent themselves), but it may also be biased. For example, a particular respondent may, for one reason or another, hold the belief that university teachers should not be expected to provide a large amount of feedback. This may lead them to interpret even only a few instances of feedback as “often” (whereas others, typically, may not).

Consequently, in such cases and with qualitative research in general, it becomes more difficult for the researcher to interpret the findings. To do so with ethical responsibility and scientific rigor might entail, for example, considering who made the statement as well as where and when they made it. The researcher would also need to take into account their own understanding of the meaning of such words as “often.” It is, therefore, important for qualitative and quantitative researchers alike to be critical not only of their data, but also of themselves in order to reduce, as best as possible, the distortion of any deductions they make through subjectivity and bias. For the same reason, the research questions were phrased in such a way as to make clear that the data collected represented the perceptions of the participants rather than scientifically precise measures. This is also mirrored in the language used throughout the presentation of the data in Chapter 4.

Indeed, the aim should be to reduce rather than eliminate bias since the latter is, according to Seale (1999), not possible: “Knowledge is always mediated by pre-existing ideas and values, whether this is acknowledged by the researchers or not” (p. 470). In fact, according to Hammond & Wellington (2013), the term “bias” is not a helpful term to

begin with since “it implies that there is a state of being unbiased” which “There is not” (p. 15).

In a similar vein, Creswell (1998) wrote that “knowledge is within the meanings people make of it; knowledge is laced with personal biases and values; knowledge is written in a personal, up-close way; and knowledge evolves, emerges, and is inextricably tied to the context in which it is studied” (p. 19). It is, essentially, an epistemological question, that is, one of “how we know what we know about […] reality and how it is that we represent that reality” (Cooper & White, 2012, p. 2).

According to Cohen, Manion & Morrison (2011), epistemology is the theory "of the nature of knowledge, its structure, organization and how we investigate knowledge and phenomena: how we know, what constitutes valid knowledge, our cognition of a phenomenon" (p. 116). More succinctly, it “refers to what we believe about how we come to know and understand the world” (Hammond & Wellington, 2013, p. 57).

There are two primary schools of thought within epistemology: positivism and interpretivism. According to Walliman (2011), positivists view the world from a materialist perspective, and hold that it is an orderly entity filled with absolutes waiting to be discovered through sensory experiences and strict experimental procedures before being translated into rules or scientific facts. Moreover, as the author explained, positivists posit that the “less measurable sciences are reducible to more measurable ones” and that even the “Social sciences can, therefore, be value free and objective” (p. 21). Quantitative research and the natural sciences in general are, thus, closely linked to positivistic thought.

Despite the above reductionist claim, there has, as Matthews & Ross (2010) noted, been a longstanding debate over the appropriateness of applying positivist thinking to social contexts. In Tracy's (2013) view, interpretivists maintain that there is no single reality waiting to be discovered, but, "Rather, both reality and knowledge are constructed and reproduced through [social processes, including] communication, interaction, and practice" (p. 40).

The ability to interpret the behaviour of the study's research participants requires researchers to identify issues from their standpoint and to understand the way in which they experience and identify with particular behaviours, events or objects (Hennink, Bailey & Hutter, 2011). This technique is rooted in the interpretive paradigm which has its roots in the previously quoted Weberian notion of *Verstehen* (Platt, 1985), denoting "true" or deep comprehension of a particular social phenomenon from the perspective of the participant. Although researchers may not be able to share precisely the same comprehension as the participants, their goal is one of "*striving toward* empathetic understanding" (emphasis added, Tracy, 2013, p. 41). Having spent a number of years living in China and engaging with individuals very similar to the study's participants will have greatly aided the researcher in meeting this goal.

3.5 Measuring Scientific Robustness

When gauging the robustness of a scientific study, its methods and of the data produced through those methods, validity and reliability are two of the most commonly employed standards (Lapan, Quartaroli & Riemer, 2012; Tracy, 2013). This section will consider these terms further in three subsections. The first provides a brief conceptualisation of the above key terms while the second examines their understanding specifically within qualitative research. The final subsection attempts to show how, in light of the

conclusion drawn in the previous subsection, the robustness of a study such as the present one should alternatively be assessed.

3.5.1 Understanding Validity and Reliability

Beginning with validity, Cohen, Manion & Morrison (2011) underscored that it is "an important key to effective research" and that, "If a piece of research is invalid, then it is worthless" (p. 179). They added that it is "the touchstone of all types of educational research" (p. 180). Validity denotes the "truthfulness" of the research findings (Lapan, Quartaroli & Riemer, 2012), that is, how accurately they reflect what they purport to reflect (Gipps, 2012). A questionnaire investigating feedback and only containing items strictly related to feedback (demographic or biological items not included), for example, would, at least superficially, indicate that it has a high level of validity.

Cohen, Manion & Morrison (2011) stated that, more recently, validity has been understood in a variety of ways such that at least twenty different types of validity have been recognised. These are usually divided into two primary types: internal and external validity. According to Walliman (2011), the former refers to "the extent to which the ideas about cause and effect are supported by the study" while the latter signifies "the extent to which findings can be generalized to populations or to other settings" (p. 104). As will be explained in more detail later in this section, external validity is largely irrelevant in this study since no claims are made as to the generalisability of its findings.

The second concept under consideration here, reliability, was defined by Tracy (2013) as "the stability and consistency of a researcher, research tool, or method over time" (p. 228). Accordingly, a reliable study is one which can be reproduced in precisely the same way, independent of who is carrying it out; a reliable research tool is one which always

functions the same way. Similarly, Hammond & Wellington (2013) stated that, while "Reliability is used in different contexts, […] in common is the sense of stable repeated measurement" (p. 131). Using the example of a questionnaire again, if the results can be replicated by other researchers using the same questionnaire, then it can be considered as reliable.

However, Matthews & Ross (2010), for example, stressed that this understanding of reliability may not be wholly suitable within the social sciences where researchers deal primarily with people rather than with inanimate objects. They pointed out that, since "Every person is an individual and different, […] no sane social science researcher would expect exactly the same result" (p. 11). Instead, the results should be similar for similar groups of people. Therefore, at least within qualitative research or studies investigating the views of people, it would not be appropriate to speak of them as simply being reliable or not.

In the following excerpt, Hammond & Wellington (2013) make the same point but also advise caution more generally when operating with the concepts of reliability and validity:

> Reliability as a concept should be treated critically. A questionnaire may be as statistically reliable as anyone can wish, but if it is not addressing one's research question it lacks validity: no matter how accurately a watch displays the time, it is never going to be the right tool for measuring temperature. In applying triangulation, contrast may be as important as convergence. There is not a state of reliability, only greater or lesser degrees (p. 131).

The two key points are, thus: neither can a method simply be either "reliable" or "unreliable" nor can a high measure of reliability guarantee a strong degree of validity. With respect to the latter measure, Hammond & Wellington (2013) argued that most decisive is the internal validity of the research. To them, this means "the logic of the

research; the clarity with which questions are formed; the fit of methodology and methods to the research questions being asked; [and] the marshalling of evidence in support of propositions" (pp. 151-152).

Upon looking more closely at the above two quotations, it can be appreciated that a study's research questions are vital in determining how reliable and valid it is. Since the participants of this study were not asked to complete the questionnaire multiple times, the reliability of their responses could not be tested. However, it is argued that a high degree of internal validity was achieved in terms of how accurately the items contained in the questionnaire reflected the research questions. Each item in the questionnaire addresses each of the study's research questions in the same order.

3.5.2 Validity and Reliability in Qualitative Research

Linking back to the previous discussion on bias, both reliability and validity are, essentially, measures of how *realitätsnah*, that is, how close to reality the research findings are. Where do reliability and validity within qualitative research stand? According to Tracy (2013), like objectivity, achieving reliability in qualitative research is unfeasible:

> traditional conceptions of reliability have little application to qualitative research, because most qualitative studies are composed of a single analysis, made at a given contextual moment in time. Because socially constructed understandings are always in process and necessarily partial, even if the study were repeated (by the same researcher, in the same manner, in the same context, and with the same participants), the context and participants would have necessarily transformed over time – through aging, learning, or moving on. Hence traditional notions of reliability used in qualitative research are not only mythical, but downright problematic […] (p. 229).

This, too, applies to the present study, for not only were the participants asked to complete the questionnaire (see Appendix 3) and be interviewed only once, but, even if this process had been repeated, any number of variables, such as those mentioned in the quotation above, may have altered their perceptions. The same goes for the researcher himself whose experiences between separate instances of data analysis may have induced him to interpret the data differently. The present study cannot, therefore, make claims for a particular extent of reliability in the traditional, positivist sense of the word. According to Stenbacka (2001), however, it does not need to since reliability, in traditional terms, concerns only quantitative measurements which played a secondary role in the present study.

What, then, of validity? Although not to the same extent as reliability, many writers agree that (positivistic) validity is also not directly applicable when researching qualitative phenomena, and that it, therefore, requires redefinition (Maxwell, 1992; Strauss & Corbin, 1998; Stenbacka, 2001; Seale, 2003). As explained by Creswell & Miller (2000), the primary reason for this lies within the understanding of the term itself: validity is affected by the researcher's perception thereof. In other words, rather than a uniform or universal concept, validity is, as argued by Winter (2000), "a contingent construct, inescapably grounded in the processes and intentions of particular research methodologies and projects" (para. 1). Thus, "qualitative researchers understand that claims are not inherently valid or invalid" (Leavy, 2011, p. 147).

It, thus, appears that reliability and validity in the traditional sense are less appropriate within qualitative research. Indeed, as argued by Strauss & Corbin (1998), "[although] the usual canons of good science have value, […] [they] require redefinition to fit the realities of qualitative research and the complexities of the social phenomena we seek to

understand" (p. 266). Watling (1995) wrote that "Reliability and validity are tools of an essentially positivist epistemology" (cited in Simco & Warin, 1997) which is used by researchers who employ experimental methods and quantitative instruments to test theoretical generalisations (Hoepfl, 1997).

3.5.3 Judging the Strength of Qualitative Research

How, then, can the strength of the present study be gauged? Typically, in the literature, the first works to be cited concerning judging the quality of qualitative enquiry are those of Lincoln & Guba (1985, 1989). They proposed using the principles of credibility, transferability, dependability, confirmability and authenticity.

According to the authors, credibility parallels internal validity and involves establishing that the findings are credible or believable from the perspective of the participants. Transferability is analogous to external validity which refers to the generalisability of the results to other samples from the same population. Dependability is akin to reliability, and signifies replicability or repeatability while accounting for the changes in the research phenomenon that inevitably occur between observations in qualitative research. Confirmability is equivalent to objectivity or the controlling for personal bias while authenticity refers to the provision of a balanced consideration of all possible or appropriate perspectives in a study. Henceforth, these terms will be used in lieu of their equivalents in quantitative research within the context of the present study due to their greater suitability within qualitatively focused investigations.

Steps were taken in accordance with the authors' suggestions to strengthen these five aspects of qualitative research. For example, the researcher had spent significant time living and working in the research setting as well as interacting with the same type of

individuals who took part in the study. Uncertainties which arose regarding the meaning of responses from the open-response questions in the questionnaire (see Appendix 3) were communicated to the relevant participants through e-mail for clarification. Such "member checking" (Cho & Trent, 2006) was also employed frequently during the interviews. Electronic notes were made throughout the research process which found use, in particular, in the construction of the interview schedule (see Appendix 4). Data were collected through multiple research methods. These measures will have bolstered the study's credibility.

The geographical, cultural and social contexts, within which the research was conducted, have been portrayed (see Chapter 1). Various facets of the research process, such as interview techniques and interpretation of the data, were periodically re-examined internally for consistency. Table 3.1 (see Section 3.7) demonstrates the chain of events in the collection of data and communication with participants. The results of the study were shared with those participants who expressed interest as well as with the host institution, thereby potentially heightening their interest and belief in the importance of the research theme. These measures will have reinforced the study's transferability, dependability, confirmability and authenticity, respectively.

Drawing on the work of Lincoln & Guba, Patton (2002) posited that the robustness of qualitative enquiry should be guided by three distinct but related elements which, together, make up its credibility:

> (1) rigorous methods for doing fieldwork that yield high-quality data that are systematically analyzed with attention to issues of credibility; (2) the credibility of the researcher which is dependent on training, experience, track record, status, and presentation of self; and (3) philosophical belief in the value of qualitative inquiry, that is, a fundamental appreciation of naturalistic inquiry, qualitative

> methods, inductive analysis, purposeful sampling, and holistic thinking (pp. 552-553).

The bar is set quite high, particularly with respect to Patton's second requirement, as this study represented only the second of its kind the researcher had ever carried out. However, although he does not enjoy a track record or status, he had taken three research methodology modules at Master level. It should be noted that Patton was also likely directing these criteria at professional researchers rather than research students.

Regarding Patton's third requirement, the researcher is strongly committed to high ethical standards, scientific rigour and professionalism throughout each stage of the research process. Moreover, he believes strongly in the merits of qualitative enquiry since, in the words of Miles (1990), it allows the researcher "to sniff the richness of the real world, see things in their contexts, track complex processes over time, and explain linkages among processes and their associated outcomes" (p. 37).

In accordance with Patton's (2002) proposed methods for enhancing the credibility of qualitative enquiry, triangulation was, as noted above, attempted as best as possible through the use of two research methods. Alternative themes, divergent patterns and rival explanations were sought, for example, through the systematic inclusion of respondents in the interview sample whose questionnaires contained a large number of contrasting and extreme responses. In addition, the researcher acknowledges his ethical and professional obligation to humility and honesty, to which the discussion of the limitations of the study (see Section 5.3) is testament.

3.6 Some China-Specific Methodological Issues

A further feature of the present study deserving attention is the context within which it was conducted. It is important to recognise and reflect on the implications of the fact that the cultural tradition whence the researcher comes (British/European or "Western") contrasts highly with that from which the study's participants come (Chinese or "Asian"). Although the present study did not involve cross-cultural research *per se* (since it did not explicitly make cultural comparisons), there inevitably existed cultural and epistemological differences between the researcher and the participants.

One issue concerns subjectivity which may be magnified if the researcher is conducting research in a cultural context different from their own. Stening & Zhang (2007) maintained that "the questions researchers ask and the theoretical constructs that underlie their work are to a very large extent a function of their own cultural background" (p. 125). This problem may be exacerbated when the research area contains several cultures or ethnicities. This is the case with China which contains 56 distinct ethnicities. Under such circumstances, researchers also need to be more familiar with the demographic characteristics of the area(s) in which they are collecting information and be aware of so-called "intracultural variations" (Kvale & Brinkmann, 2009). This issue, however, would likely have been less relevant within the research area of the present study since it is dominated by the same ethnicity (Han) which, in 2010, accounted for over 91% of the total Chinese population (National Bureau of Statistics of China, 2011).

The issue of subjectivity in intercultural or interlingual research may be further amplified when a study's focus revolves around a particular construct. Not all concepts may be directly translatable and, for some, an equivalent term may not even exist. Levine, Park & Kim (2007) emphasised:

> Colloquially, researchers need to compare apples and apples, not apples and oranges. More formally, the conceptual definitions of the constructs under investigation need to be constant, general, and equally valid in each of the cultures compared (p. 209).

Despite being a sound requirement in theory, it may be difficult to achieve depending on the construct being examined. In fact, such concurrence may not even exist within one culture. The major concept involved in the present study was “feedback.” The discussion in Section 2.6 has shown that some consensus on what broadly constitutes feedback exists. Whether such consensus exists in the Chinese literature is not known. Feedback does, however, appear to be an established term in English articles on educational assessment published by Chinese scholars, typically within discussions on formative assessment (e.g., Wang, 1996; Cheng, Rogers & Hu, 2004; Wei & Chen, 2004; Wang & Lê, 2006; Shing & King Fai, 2007; Wang, 2007; Wang, 2008; Qu & Yang, 2010; Wang, 2010; Wei, 2010; Zhan & Wan, 2010; Zhao, 2010; Wei, 2011; Zhu, 2012; Tang, 2013). It was, therefore, deemed acceptable to use this term in the Chinese context.

Despite this, feedback is, according to the researcher’s experience, not a term all the participants will have been familiar with in English. Therefore, so as to make an otherwise abstract idea more concrete and to minimise misinterpretation, in addition to the equivalent word in Chinese being given, the term was also defined and exemplified in the covering letter (see Appendix 1). During the piloting of the questionnaire (see Appendix 3) and its accompanying documents, no students noted any difficulty understanding the concept or its accompanying explanation. In addition, none pointed out an incongruence between the Chinese term and the English term and exemplification. Some extent of conceptual validity, thus, appears to have been achieved.

A second problem that may arise when conducting research in China, in particular through a questionnaire, is response set bias. According to Stening & Zhang (2007), and as confirmed by a number of authors cited by them, there appears to be a tendency for Chinese respondents to select neutral responses. They pointed out that this likely results from a "deep-seated Confucian virtue, the 'doctrine of the mean', placing a high value on moderation and avoiding extremes" (p. 130).

A consequence of this canon in terms of using a questionnaire, where (many of) the respondents are (likely to be) Chinese or from a Confucian-heritage setting, is that neutral responses should, where possible, be avoided. On the other hand, the neutral option may, some cases, be the most accurate response. Not offering a neutral option may, therefore, significantly lower the accuracy of the data. Thus, it was decided to include a neutral response but, where possible, to place it in the last column on the right-hand side so as to prevent it from being spatially in the middle.

Although writing within management theory, Roy, Walters & Luk (2001) offered a number of helpful suggestions on constructing and implementing questionnaires within survey research in China. With respect to tackling response set bias, they recommended using an even-numbered rating scale. Despite this, a five-point scale was chosen for the questionnaire in this study since four points were perceived as lacking accuracy and six or more as being burdensome due to a much higher degree of precision. Another point regards metric equivalence which the authors point out is important though difficult to determine in cross-cultural research due to difficulties in the translatability of points along a scale. However, this was not an issue in this study since a semantic differential scale was always used.

Another issue pointed out by Roy, Walters & Luk, which the researcher has also found to be particularly salient in all dealings involving interpersonal communication in China (and, hence, is particularly relevant to interviews), is what they termed “social acquiescence.” This refers more specifically to the matter of “face” which is a prominent communicative feature of particularly Chinese societies (Hwang, Ang & Francesco, 2002). The concept of “losing face” refers to humiliation or being made to look foolish, and, according to Hofstede (2003), should be avoided at any cost.

Roy, Walters & Luk (2001) pointed out that, in order to “save face,” respondents may pretend to come across as more knowledgeable on a topic than they actually are. Alternatively, they may report what they believe the researcher expects to hear in order to enhance the interviewer’s face. Interviewees may also withhold information or report events inaccurately in an effort to avoid jeopardising social relationships with third parties.

This first of the above concerns will not have been relevant since interviewees were not asked any questions which required displaying any breadth of knowledge on feedback, rather just of their own experiences of teacher feedback. The second issue will also have been of little relevance since none of the interviewees were any longer enrolled in the researcher’s courses and, thus, stood to gain little from telling him what they believed he wanted to hear. The third matter is more significant since some interviewees may have feared putting their instructors in a bad light through reporting, for example, having received unhelpful or insufficient feedback from them. However, it is also not expected that this will have occurred often since all interviewees were guaranteed anonymity within the discussion of the results and were also not required to specify the names of instructors.

Stening & Zhang (2007) made a number of other recommendations when conducting research in China. Those that are particularly relevant here are that: face-to-face interviews are more reliable and more likely to be successful than telephone interviews or postal questionnaires; participation should be voluntary (that is, not given because of pressure from superiors) and anonymity should be assured; and participation should be sought through and sanctioned by senior intermediaries (so as not to step over hierarchical boundaries). All three recommendations emanate from the nature of Chinese society and prevailing customs guiding interpersonal communication.

The authors' first guideline was met fully, as all interviews were conducted face-to-face. Although permission was granted by the relevant department, no higher authorities except the researcher and a small number of his colleagues were directly involved in soliciting participants. Moreover, great emphasis was placed on the voluntary nature of participation and on the absence of any impact on students' course marks from both participation and non-participation. Stening & Zhang's second guideline – also required by the British Educational Research Association (2011) – was, therefore, also satisfied. Their third guideline was adhered to through an initial meeting with and authorisation from heads of department. Related ethical issues are discussed further in Section 3.11.

A further important matter that deserves consideration within research in China, or with Chinese participants, concerns insider versus outsider research. This may apply to any context, but perhaps even more so to that where the researcher is not (considered as) a national or cultural affiliate. It is a topic that is discussed extensively in the literature, for instance by Mercer (2007) who raised awareness of some issues inherent in so-called insider research. She deliberately avoided attaching a specific definition to the term "insider" since she believed that whether a particular researcher is an insider or an

outsider is context-dependent; moreover, it is, to some extent, possible simultaneously to be both.

Despite making this caveat, Mercer employed the term “insiderness” to denote a degree of familiarity with the research venue as well as the researcher having one or several characteristics in common with the individual(s) native to that context. One of the major challenges posed by insiderness or “familiarity” which Mercer drew attention to is short-sightedness and reduced objectivity. In other words, insider researchers may take their access privileges for granted and may also assume that their views are more pervasive than is actually the case. Heightened familiarity may (even more so) produce behaviour on the part of those being researched which either was not there in the first place or which does not equate to the *status quo*.

According to Mercer’s understanding of the term, the researcher of the present study is indeed an insider since, as reiterated above, he has spent a considerable amount of time in China, including several years in the research location. Despite this, it is doubtful that he was, in fact, perceived as an insider due to apparent differences not only in physical appearance, but also in language and certain mannerisms. He argues that he possesses a sufficient familiarity with the cultural context within which the present study was set to act in accordance with local etiquette and proceed with the necessary discretion.

Finally, there exists one possible challenge of a linguistic nature: all participants were non-native English speakers. Lindsay (2010) suggested that, where this is the case, “communication in the most appropriate language […] must be considered” (p. 119). Since the researcher’s Mandarin Chinese skills are not proficient enough, it was decided

to conduct all communication in English. Due to time and financial constraints, the use of translation services and the hiring of an interpreter were not feasible. The researcher does, however, have several years' experience communicating with Chinese nationals with varying degrees of English proficiency both in and outside of China. Care was taken to phrase all questions and terms used during the interviews as well as in the questionnaire so as to be more easily understandable by non-native English speakers.

3.7 The Study at a Glance

As a transition to the more concrete methodological aspects of the current study, an overview of the major stages involved in its preparation and the data collection process is provided. Table 3.1 summarises the major stages, their approximate duration and the dates during which they were carried out. More specific descriptive meta-data pertaining to the questionnaire and interview are discussed in Section 3.12.3.

Step No.	Step Description	Period	Timeframe
1	Identification of study focus, initial search and review of literature, and defining of research questions (see Section 2.10)	8 weeks	Sept. 2012 – Apr. 2013
2	Meetings with department heads, explanation of research intentions, and obtaining of approval to conduct study at host institution	2 weeks	Oct. 2012
3	Obtaining of ethical approval	1 week	May 2013
4	Drafting and redrafting of questionnaire, covering letter and participant rights letter	3 weeks	June – Aug. 2013
5	Selection and contacting of pilot study participants	1 week	Sept. 2013
6	Piloting of questionnaire, covering letter and participant rights letter	1 week	Sept. 2013
7	Editing and release of final version, covering letter (see Appendix 1), participant rights letter (see Appendix 2) and questionnaire (see Appendix 3) as Microsoft Word files.	1 week	Sept. 2013

8	Receipt and storing of completed questionnaires (*n* = 232), and tabulation and processing of data in Microsoft Excel	20 weeks	Sept. 2013 – Feb. 2014
9	Analysis and coding of questionnaire data in Microsoft Word (see Appendix 6) and NVivo	2 weeks	Feb. 2014
10	Drafting and redrafting of generic interview schedule	2 weeks	Feb. 2014
11	Selection and contacting of pilot study participants	1 week	Mar. 2014
12	Piloting of generic interview schedule	1 week	Mar. 2014
13	Editing of final version of generic interview schedule (see Appendix 4)	1 week	Mar. 2014
14	Selection of and contacting of interviewees	1 week	Mar. 2014
15	Conducting of face-to-face, one-on-one interviews (*n* = 30)	8 weeks	Apr. – May 2014
16	Transcription of interview recordings in Microsoft Word	5 weeks	June – July 2014
17	Examination of electronically transcribed interview recordings (see Appendix 7)	3 weeks	July 2014
18	Search for recent literature	2 weeks	Aug. 2014; Mar. 2015

Table 3.1: Data collection timetable

The periods are, in some cases, approximate and aggregate, and some of the steps overlapped. Redrafting the questionnaire, covering letter, participant rights letter and generic interview schedule prior to their piloting was done according to feedback received from the researcher's supervisor. Editing the final version of the questionnaire, covering letter, participant rights letter and generic interview schedule was done according to feedback received from the pilot participants. The piloting process is described in greater detail in the following section.

3.8 Pilot Study

Piloting refers to testing data collection tools on individuals who are similar to those who will be included in the main study (Lapan, Quartaroli & Riemer, 2012). Cohen, Manion &

Morrison (2011) stated that a pilot study's principal functions are "to increase the reliability, validity and practicability" of the method (p. 402). May (2011) repeatedly underscored the importance of piloting research instruments in social research since it allows researchers to "assess both individual questions and how the measurement tool functions as a coherent whole in the field" (p. 107).

May's point applies, in particular, to questionnaires but also to structured or semi-structured interview schedules that contain closed items where the researcher will seek to compare datasets and draw conclusions from them reliably and validly. Bryman (2012) pointed out that piloting is all the more crucial in the case of self-completion questionnaires since the researcher is not present to resolve any confusion on the part of respondents, potentially resulting in irrelevant or unusable data and considerable wastage. With these recommendations and warnings in mind, both research instruments used for the present study as well as their accompanying letters were carefully piloted.

The participant rights letter, questionnaire and covering letter were piloted with a total of 17 undergraduate students (8 females and 9 males). Pilot study participants were solicited through an announcement made in each of the researcher's own classes, from each of which two or three of the first to volunteer were chosen. The sampling strategy employed for finding questionnaire respondents and selecting interviewees is detailed in the following section.

Before piloting, all English terms translated into Chinese contained in these documents were checked for accuracy by a certified translator. Based on the comments received from the pilot study participants, only a small number of minor changes needed to be made to the questionnaire, indicating that the participants were satisfied with its

language, content and layout. Most noteworthy and as explained in Section 2.6, the phrase “self-efficacy” was replaced by “self-confidence” in the questionnaire (see Appendix 3). A definition of these and other phrases was provided beneath the corresponding questionnaire item for purposes of clarification. Finally, the instruction to use crosses when selecting options was changed to using checkmarks, and the term “gender” was used instead of “sex.”

The generic interview schedule was piloted with 5 undergraduate students (4 females and 1 male). Only one but an important change that was made after piloting was to add the first question which asks the interviewees to state their subject of study. This was done in light of comments received from several of the students piloting the interview schedule. They suggested that this aspect be taken into account due to potentially significant differences in how students from different academic backgrounds receive and experience feedback. A possible correlation in this connection was investigated, and the results are presented and discussed in Section 4.4.

Originally considered for inclusion in the questionnaire was an item investigating on which levels students receive feedback as a parallel to Hattie & Timperley’s (2007) taxonomy discussed in Section 2.6. However, comments received from those students piloting the questionnaire showed that the level of detail and language necessary for doing so was too high. Therefore, as discussed in the same section, two of the writers’ feedback levels were explored through the interview schedule and the other two not at all due to an even higher level of complexity and greater difficulty in explaining their meaning.

3.9 Sampling

3.9.1 Sampling Methods

According to Cohen, Manion & Morrison (2011), there are two general approaches to sampling: probabilistic (random) and non-probabilistic (purposive). In the latter case, the chances of members of the broader population being chosen for the sample are known whereas, in the former, they are not. The authors explained that, in a probability sample, each member of the broader population has an equal chance of being represented in the sample; inclusion as well as exclusion are only a matter of chance. In a non-probability sample, on the other hand, some members may have a greater chance of being included or excluded than others, or some may certainly be included while others may certainly be excluded. This study employed a non-probabilistic approach.

As Cohen, Manion & Morrison pointed out, non-probabilistic sampling is typically used in small-scale qualitative research where a particular group of individuals is targeted and, perhaps most crucially, where no attempt at generalising to the broader population is made. The core reason this approach was employed in this study was only those students whose English was at a high enough level could take part. Neither the researcher's level of Chinese nor the resources available to him were sufficient to allow him to produce the questionnaire (see Appendix 3) or conduct the interviews in Chinese, or to translate any data collected into English. A further reason was the logistical impossibility of directly accessing and taking a sample from the entire student body (over 40,000). This would have necessitated consulting with or involving too many department heads, staff members and other intermediaries for it to be practical. The sample is described in more detail in the following subsection.

There are several types of non-probabilistic sampling, and three were employed in the present study: convenience, snowball and purposive sampling. Cohen, Manion & Morrison defined convenience sampling as the selection of those individuals who are nearest or happen to be available, and that this process is continued until the requisite total number of participants has been reached. The first two sampling methods were employed to gather questionnaire respondents.

Convenience sampling was achieved through rendering the questionnaire electronically as a Microsoft Word document and uploading it to the web pages of all of the researchers' courses and of some of those of colleagues. These web pages are located on the university's internal server and are accessible by all students, teachers and administrative staff at the host institution. Several fellow teachers (only those who taught students who met the language requirement) were involved in its dissemination through this method. Accompanying the document was a message containing a brief description of the study and an invitation to take part by downloading the attached questionnaire and returning it by e-mail to the researcher. Participants were also given the option of printing out the questionnaire and submitting it in person. Alternatively and in an effort to offer anonymity, participants were invited to place their questionnaire into the researcher's on-campus post box. Further steps taken to offer anonymity are discussed in Section 3.11.

Use was also made of snowball sampling which, according to Cohen, Manion & Morrison, is "often pre-eminent in qualitative research" (p. 159). This method involves the identification of a small number of suitable participants who, as informants, identify and invite others who meet the same eligibility requirements who, in turn, do the same. In this study, students were not approached individually with the expectation that this would lead to an exponential growth in the overall number of participants. However,

those from whom a questionnaire was received electronically were invited to introduce the study to other students (again, only to those who met the language requirement).

The key feature of the chosen sampling strategy was, therefore, that it relied on exposure to a large number of potential participants simultaneously as well as, though to a lesser extent, on word of mouth. One weakness inherent in this method is that there was, with the exception of making brief announcements in class, little opportunity for encouraging students individually to take part or for participants to respond to all the questions. The majority of the onus in this regard fell on the covering letter (see Appendix 1). Thus, great care was taken to formulate it in such a manner to be as inviting and persuasive as possible without being forceful. Where possible, a brief summary of the study and an invitation to take part were also given orally to potential participants in class.

Another drawback related specifically to snowball sampling is that, as Cohen, Manion & Morrison cautioned, it is more prone to bias since those who were included in the sample through this method may be influenced heavily by those who were used initially as informants. This may arise from the fact that, likely being acquaintances, an informant and another participant may share similar views or come from similar backgrounds. Depending on how long such a chain of participants is, the likelihood that the data collected from these individuals is biased or skewed is much higher. However, it appears only a small number of participants were found this way since the majority of respondents confirmed in their message containing the complete questionnaire that they had seen the notification on the website. Six students had taken it upon themselves to invite other students to take part in the study, resulting in a total of 15 additional questionnaires being returned in this manner.

The study strived to collect completed questionnaires from as many eligible students as possible, not in order to be able to generalise to the rest of the population but, primarily, to increase the likelihood of finding vocal respondents willing to go into greater depth on the phenomenon under investigation. It was, furthermore, not expected that such a high percentage of respondents would provide answers to the open-response questions in the questionnaire nor that so many would agree to an interview. It should also be noted that, in the case of the questionnaire respondents, neither the word "sampling" nor the word "selection" are appropriate since participants were not sampled or selected as such; more aptly, they were "captured."

In contrast, interviewees were deliberately chosen. Cohen, Manion & Morrison referred to this method as purposive sampling which refers to researchers including cases in their sample "on the basis of their judgement of their typicality or possession of the particular characteristics being sought" (p. 156). Brinkmann (2013) labelled this method as an "information-oriented approach" which typically aims for in-depth analysis rather than generalising to the rest of the population. Citing Flyvbjerg (2006), Brinkmann (2013) explained that, in an information-oriented approach, researchers may look for: "extreme cases" which may uncover the phenomenon in its "purest form;" "maximum variation cases" from which researchers may glean the significance of highly divergent cases; "critical cases" from which knowledge may be obtained to make deductions and falsifications (e.g., "If the responses provided by respondents X, Y and Z were similar and respondents X and Y noted phenomenon A, then respondent Z might also be expected to"); and "paradigmatic cases" which may allow researchers to establish or confirm shared characteristics (pp. 57-58).

Use was made, in particular, of the first two cases. Regarding extreme examples, the underlying question was “Why did this student provide such a high number of extreme responses?” or “Are there are any idiosyncratic factors that can account for their extreme responses?” Concerning maximum variation cases, the essential question was “What is different about students who provided extreme responses at one end of the spectrum and those who provided extreme responses at the other end of the spectrum?” or “How do these students contrast?” Critical cases and paradigmatic cases were not isolated intentionally but were included nonetheless.

In order to achieve greater authenticity, all respondents who were postgraduate students were included in the interview sample. This enabled the incorporation of the perspectives of postgraduate as well as undergraduate students. In addition, those who provided unique or copious comments were also included since it was expected that they might be more willing or able to provide greater insight. Moreover, efforts were made to include an equal number of males and females, though this could not be controlled entirely due to giving priority to including examples of the four cases described above.

Brinkmann admitted that “Sometimes qualitative interviewers do not have the luxury of choosing a sampling strategy, but must stick to the respondents that they are able to recruit” (p. 58). In the case of this study, this applied, in particular, to the selection of interviewees since around half of the respondents did not offer to be interviewed. This, however, was not a grave concern since the aim was not to generalise to the rest of the population but to more closely examine those cases which held the greatest promise of shedding light on and aiding in the understanding of the research phenomenon.

3.9.2 Sample Population

Participation was made open to all of the researchers' Chinese undergraduate students as well as to a number of his colleagues' Chinese undergraduate and postgraduate students with an English level above intermediate at one National Key University from all departments (Note: "intermediate" denotes that individuals are able to engage in at least basic conversation and express themselves intelligibly, with adequate fluency, and with sufficient grammatical accuracy and lexical range). Due to the sampling methods used, students may have been drawn from any number of different faculties from all over the university. Students were not discriminated on the basis of ethnicity, gender or age.

A total of 232 questionnaires were collected, of which 230 were received electronically and 2 submitted in person. None were placed in the researcher's post box. Male participants were in the slight majority with 127 (54.7%) while 105 questionnaires (45.3%) were returned by female participants. The vast majority of participants were undergraduates, totalling 223 (96.1%), while postgraduates accounted for 9 (3.9%). The low number of graduate participants can be explained by only a small number of such students having been able to be accessed and only through second parties.

A total of 30 interviews were conducted, 16 with male students and 14 with female students; 26 with undergraduate students and 4 with graduate students. The average duration was 25 minutes and 11 seconds, with 12 minutes and 13 seconds being the shortest and 52 minutes and 17 seconds the longest. All interviewees consented to the interview being recorded and none requested the interview to be paused or halted, or to skip any questions.

3.10 Data Analysis

The analysis and presentation of the quantitative data gathered through the closed items in the questionnaire employed primarily descriptive statistics. All such data was processed in Microsoft Excel. With the exception of one individual enquiry (see Section 4.4), the approach used in analysing, categorising and interpreting the qualitative data collected through the open-response questions in the questionnaire and the interviews was inductive, a hallmark of grounded theory.

Developed initially by Glaser & Strauss (1967), “grounded theory probably represents the most influential general strategy for conducting qualitative data analysis” (Bryman, 2012, p. 575). According to Guest, MacQueen & Namey (2012), “grounded theory is a set of inductive and iterative techniques designed to identify categories and concepts within text that are then linked into formal theoretical models” (p. 12). In other words, a grounded approach sets itself from deductive methods in that, instead of testing for certain hypotheses, it works backwards to formulate hypotheses that are *grounded* in the data themselves; it is exploratory rather than confirmatory. This reflects the nature of this study since its primary aim was to uncover the views of the participants and the possible reasons for these views based on the comments collected from them.

Developed by writers on grounded theory and representing the starting point for most forms of qualitative data analysis is coding (Bryman, 2012). According to Lapan, Quartaroli & Riemer (2012), “Coding involves the classification of elements in text data into categories that are related to the study topic and are useful in analysis” (p. 98). Cohen, Manion & Morrison (2011) differentiated between three core types of coding. The one that applies most to the present study is open coding which “involves exploring the data and identifying units of analysis to code for meanings, feelings, actions, events […]”

(p. 600). Through this process, categories and subcategories are created, and possible connections are made between them.

The codes into which fragments of data are grouped during the coding process are allocated particular units of meaning (Walliman, 2011). Each code is given a label "that simultaneously categorizes, summarizes, and accounts for each piece of data" (Charmaz, 2014, p. 111). These segments of data are sometimes termed "themes," the identification or generation of which is commonly called "thematic analysis." Despite not being tied to a specific discipline or epistemology, and although not a separate method *per se* (Boyatzis, 1998), thematic analysis is one of the most common approaches to qualitative data analysis (Bryman, 2012). It also represents the primary method of data analysis employed in the present study.

Since the word "theme" is a key concept within the analysis and discussion of the present study's findings, it bears briefly explaining what it means and how examples of its kind can be identified. Boyatzis (1998) provided a definition which answers both these questions:

> A theme is a pattern found in the information that at minimum describes and organizes the possible observations and at maximum interprets aspects of the phenomenon. A theme may be identified at the manifest level (directly observable in the information) or at the latent level (underlying the phenomenon). The themes may be initially generated inductively from the raw information or generated deductively from theory or prior research (pp. vi-vii).

This definition can be used as a framework to describe the process through which themes were generated from the data gathered through this study. Themes were described and organised; they were also interpreted as best as possible through triangulation between datasets collected through the two research instruments. Since

the coding process rested primarily on keywords, the identification of themes can be said to have taken place chiefly on the manifest level since this information was directly observable.

This is not to say that information which appears “manifest” is not susceptible to misinterpretation; threats to both credibility and confirmability are omnipresent. Nonetheless, the researcher attempted to limit this threat by, for example, grouping only those responses within one theme that contained keywords which were synonymous, stemmed or inflected forms of each other according to NVivo. Finally, through this process and as already discussed earlier, themes were, for the most part, generated inductively.

As indicated in Table 3.1 (see Section 3.7), two strategies were employed in conjunction to code and identify themes in the data. The first involved copying and pasting the open-response questions from the questionnaire into a Microsoft Word document – one per set. Participants’ transcribed responses to the interview questions were also divided into sets in order to allow for more convenient comparison. Hammond & Wellington (2013) confirmed that aggregating and analysing data from open-response items in such a way is a frequently used method by researchers in the case of surveys as well as structured and semi-structured interviews. Each set was read multiple times; once to make minor corrections, a second time to internalise students’ responses, and additional times to identify themes. This process was repeated after seven days to enhance the dependability and confirmability of the findings or, more specifically, to ensure that potential themes or relevant responses had not been overlooked or erroneously included. Indeed, through this way, a small number of additional responses were included in some of the themes.

With respect to making corrections, only a very small number of spelling mistakes and other obvious errors (some of which are typical of Chinese) – such as “write” (right), “knowledges” (knowledge), “shortages” (shortcomings), “revisement” (revision) and “hobbit” (habit) – were corrected. This was done for the sake of these words being identified (correctly) by NVivo. However, in order to maintain the highest possible level of authenticity, grammatical inaccuracies contained in participants’ responses quoted directly in this thesis were not corrected.

Two methods were used during the isolation of themes through the first strategy. One was to electronically highlight recurrent keywords and phrases. A different colour was used for each theme. A second method was to make notes – also electronically – to accompany already established as well as potential other themes. Glaser (1998) referred to these notes as “theoretical memos” and argued these to be a key component of open coding. These two methods were also used in the examination and comparison of interview responses. Extracts showing the use of colour highlighting and electronic note-making with the questionnaire and interview data are provided in Appendices 6 and 7, respectively. Colour highlighting played a more minor role in the case of the interview responses since these served primarily to explore more deeply those themes which had already been identified in the questionnaire data.

The second strategy to help reveal possible themes entailed the utilisation of NVivo. Use was made, in particular, of the Word Frequency and Text Search functions to establish the frequency of concepts and ideas within and across responses. Ryan & Bernard (2003) recommended looking for the recurrence of terms and topics as one method of identifying themes. Bryman (2012) also pointed out that repetition is likely one of the most common criteria for designating a particular code as a theme, but warned that, in

isolation, repetition is insufficient for doing so. He underlined that “it must be relevant to the investigation’s research questions or research focus” (p. 581). This was warranted through continuous cross-referencing of potential themes with the research questions. In addition, only those items which were included in the questionnaire or aspects which arose from students’ responses to the open-response questions which had a dataset coverage of at least 5% were included in the discussion of the study’s findings (see Chapter 4). This was done primarily due to space constraints.

It needs to be emphasised that NVivo was not used to dictate but rather only to assist as an additional tool to triangulate themes identified through the first strategy or to uncover possible new ones. This is, in fact, how Hammond & Wellington (2013) argued the role of computers should be understood, namely as aids in the coding process but not as the coders themselves. In her juxtaposition of constructivist and objectivist grounded theory, and criticism of the latter, Charmaz (2000) argued that computer-assisted software programmes fragment data superficially which isolates data from “our constructions and interpretations” (p. 521). Ryan & Bernard (2003) remarked that, once the researcher has “a feel for the themes and the relations among them, we see no reason to struggle bravely on without a computer;” even in the case of hundreds of interview transcripts, “there is no substitute for following hunches and intuitions in looking for themes to code in texts” (p. 94).

A further motivation for the greater reliance on the first strategy and also one of the major challenges in the transcription and coding of participants’ responses was a number of linguistic issues, including grammatical inaccuracy, wrong word choice and, in the case of the interview recordings, the frequent occurrence of unfinished sentences and pauses. Since the data was transcribed *ad verbatim* (with the exception of the

corrections described above) and due to the researcher's intimate familiarity with how Chinese communicate in English, it was felt that the first strategy, though inferior in terms of dependability, would be less prone to misinterpreting the data and, therefore, offer greater confirmability.

3.11 Ethics

Ethical approval (see Appendix 5) was sought from and granted by the University of Warwick Institute of Education (now Centre for Education Studies), and the University of Warwick's (2013) *Research Code of Practice* was read prior to the collection of any data. The British Educational Research Association's (2011) *Ethical Guidelines for Educational Research* was also studied and used as a yardstick to ensure high ethical standards whilst collecting data.

The ethical implications of this study can be considered to have been reasonably minimal since the research topic was largely independent of socio-political themes and also because the research questions were not of intrusive or personal nature. In addition, all participants were adults and voluntary informed consent was obtained beforehand. This was facilitated principally through the participant rights letter (see Appendix 2) which was referred to again as a physical copy at the outset of the interview.

Interviewees were reminded before beginning the interview that they could skip any questions they felt unable or uncomfortable answering. The rights to terminate the interview prematurely and to withdraw from the study entirely were also furnished. Permission for the interviews to be recorded was requested explicitly but not insistently, and it was explained that the reason for doing so was to allow for transcription and closer examination of the conversations. Interviewees were assured that only the

interviewer would have access to the recordings and transcripts, that these as well as any other information provided by them would be stored securely, and that they would be deleted upon completion of the study. These rights will have further bolstered voluntary informed consent.

Due to the necessity of reporting the findings of this study, confidentiality could not be offered and complete non-traceability also could not be guaranteed. However, great care was taken to paraphrase or not to reproduce any oral or written comments from participants that contained information which could be traced back to them. Anonymity was rendered through employing identity codes when referring to individual questionnaires and interviewees. All material quoted from participants is referred to either with “Q” or “I” which denote questionnaire and interview responses, respectively. Both labels were used in cases where references were made to the questionnaire and the interview conducted with the same participant. Proceeding this designation is either “U” or “P” which signifies whether the statement was made by an undergraduate or postgraduate student, respectively. Following this is the number which corresponds to the order in which the participant’s questionnaire was received. Any mentions of the name of the university where the research was conducted in quoted material from participants as well as in the appendices have been redacted or anonymised through the usage of the acronym “HI,” signifying “host institution.”

While there will have been little potential for the study to improve the personal situation of the participants, it is, for the reasons noted above, very unlikely that the study or any questions posed to the participants will have in any way harmed them. With the exception of some instances where interviewees struggled to find the appropriate language to express themselves, no other signs of distress were detected. To allow

interviewees to be better prepared and feel more assured during the interview, they were provided with a copy of the interview questions (approximately one week prior). Several of them brought an annotated copy of the interview questions with them to the interview. Hence, the principle of non-maleficence will also have been largely observed.

3.12 Research Instruments

3.12.1 The Questionnaire

The questionnaire was chosen to spearhead the information gathering due to the convenience of its dissemination and publicising electronically through the Internet. In addition, as Verma & Mallick (1999) stated, questionnaires (which typically, though not always, precede interviews) may elicit unexpected insights around which the interview should be fitted. The following subsections outline the major advantages and disadvantages of employing questionnaires for research purposes. Measures adopted to counteract the disadvantages as best as possible are also explained.

3.12.1.1 Benefits

A number of characteristics of questionnaires speak in favour of their usage. Arksey & Knight (1999) pointed out that, compared to other methods, they offer greater potential for anonymity; can be completed at a time and in a place convenient for the respondent; and elicit greater willingness from respondents to take part because of less intrusiveness from the researcher. All these points endorse the usage of questionnaires in terms of ethics.

The authors also called attention to the convenience of questionnaires: they produce results quickly and they can be completed in the absence of the researcher, thereby reducing potential sources of bias. A further advantage associated with the second point

is that questionnaires do not have to contend with the potential problem of respondents not being available at the time they are distributed since they can be completed and returned subsequently. Moreover, they are cost-effective due to being able to be disseminated to a large group of potential participants.

In addition, as argued by Sarantakos (2005), questionnaires may offer a measure with greater stability, consistency and uniformity, which bolsters the reliability and validity of the findings. They also provide respondents with the opportunity more carefully to consider their answers and viewpoints due to being in written form. This point may be particularly important in the case of respondents who are required to answer questions in a foreign language, which may require greater time expenditure.

3.12.1.2 Drawbacks

Despite this considerable list of advantages, questionnaires are not without their weaknesses which must be considered and, where possible, offset. One, as highlighted by Gillham (2007), is that, since they are typically completed in the absence of the researcher, they do not allow for real-time deeper probing, follow-up questions or for the clarification of their contents. The second point was addressed by inviting participants to contact the researcher by e-mail or phone in the case of uncertainties about any of the items. No such enquiries were received. In addition, as noted in Section 3.8, the pilot study participants had raised few questions about the language, content or layout of the questionnaire. This suggests that respondents in the main study would not likely have required clarification of any facet of the questionnaire.

Gray (2004) warned that questionnaires – perhaps with the exception of web-based ones and those completed in the presence of the researcher – do not offer any insight as

to whether the intended order of the items was followed. Doing so might be important, for example, if one item can influence the answer to another. Following the order of items in this study was, however, not pertinent since the items were few in number and largely independent of each other.

Sarantakos (2005) also cautioned that, if questionnaires are completed in the absence of the researcher, it cannot be ensured that they were completed by the appropriate individuals or, if relevant, in the appropriate conditions. In addition, questionnaires may be returned only partially completed. Respondents may also complete them more hastily and without careful consideration.

Regarding the first of these issues, there would have been little motivation for someone to complete a questionnaire on behalf of another person since there is no obvious way in which they could profit from doing so. No particular conditions were necessary within which to complete the questionnaire, and so this problem was not relevant.

With respect to the third challenge, no partially completed questionnaires were returned, although it cannot be known how many may have been partially completed before being abandoned. Three questionnaires lacked one response within one of the matrix questions, which, as it turned out, occurred unintentionally. The missing information was recovered through contact with the relevant participants through e-mail.

The fourth problem is not likely to have been a common occurrence in this study since participation was voluntary and since it is likely that only those who were interested in sharing their views would have taken the time to participate. In addition, as reported in

Section 3.12.3, a large number of respondents made the effort of offering answers to the three open-response questions, demonstrating genuine interest in the research subject.

Sarantakos further underscored that collecting data through questionnaires does not offer the researcher the opportunity to collect supplementary information while they are being completed. This was one of the motivations for combining this method with an interview so that such information could be captured at a later time. Each of the interviewees was reminded of their previous responses in order to prompt views or experiences they may have had in mind at the time they completed the questionnaire.

Inherent in questionnaires may also be a number of linguistic as well as technical concerns. For example, Gillham (2007) pointed out that not all respondents may be literate. This, however, would not have been a barrier in this study since all participants were university students and will, therefore, have been literate. The same applies to computer literacy: all students will have been capable of downloading, processing, uploading and transmitting electronic documents since these skills are necessary for their courses. It was also known that all students had access to a computer and to the Internet.

It is acknowledged that there will likely have been some variation in English proficiency between the respondents since participation was made open to students from any level above the intermediate level. For this reason, the lexical difficulty was pitched at a lower, but still appropriate, level to ensure that all respondents will have been able to understand and respond to the questions. A high density of academic or technical expressions would have been a source of bias since only those participants whose English proficiency was at a standard high enough could have understood them.

Although typically more relevant to web-based questionnaires, a further potential problem arising from electronic questionnaires is technical malfunctions from server errors, system crashes, incompatibility issues and other computer problems, which may cause respondents to abandon the questionnaire (Fox, Murray & Warm, 2003; Parsons, 2007). These problems were acknowledged but seen as an inevitable hazard to the convenient and low-cost distribution and collection of questionnaires through an electronic medium.

3.12.2 The Interview

The interviews served primarily to gain deeper, more qualitative information as well as to build on that collected through the questionnaires. Yin (2011) stated that semi-structured or qualitative interviews, which include open-response questions, are likely "the dominant mode of interviewing in qualitative research" (p. 134). The following section outlines the major advantages and disadvantages of employing interviews for research purposes. Again, measures adopted to counteract the disadvantages as best as possible are explained.

3.12.2.1 Benefits

Brinkmann (2013) averred that, "when one wants to know how an individual experiences some phenomenon, interviewing has a certain primacy among the different methods" (p. 47). There are a number of positive attributes inherent in interviews discussed in the literature. According to Arksey & Knight (1999), they generally enjoy a high attendance rate given the direct communication that normally occurs between the researcher and the participant. For the same reason, they also offer greater control over the environment or location and the time during which they are conducted. These aspects can normally be specified by the interviewer or negotiated with the interviewees.

Particularly in the case of structured and semi-structured interviews, this method also offers the interviewer greater control over the order of the questions.

A large number of benefits of interviews are directly related to the proximity of the participant to the researcher. For example, Gray (2004) pointed out that interviews enable the interviewer to have certainty over the source of the information. They also provide an opportunity for questions and misunderstandings to be addressed immediately, thereby increasing the likelihood that more questions will be answered or answered more completely which, in turn, strengthens the validity of the findings. Interviews are also less wearisome and less monotonous for participants since there is two-way interaction and cooperation. Moreover, face-to-face interviews are, as explained in Section 3.6, more appropriate in the Chinese context.

Sarantakos (2005) also called attention to the opportunity that interviews present to record spontaneous answers (which may yield new and valuable insights) and pose more complex questions since the researcher is present to provide any necessary assistance in answering them. On a similar and more general note, it is argued here that interviews have the potential for collecting data that is much deeper than a number of other methods since more extended and detailed responses can be elicited from the interviewees. This information, in turn, can help researchers gain a better greater understanding of the phenomenon under investigation and enable them to make a stronger case for particular deductions. This may eventually lead to instilling a greater degree of confidence from the target audience in a researcher's evidence and conclusions.

3.12.2.2 Drawbacks

As with the questionnaire, interviews have drawbacks which should be reflected on before they are employed in the field. Perhaps the most obvious of these is that they, as also argued by Arksey & Knight (1999), are often more costly and time-consuming and, therefore, less convenient. Face-to-face interviews naturally require the physical presence of the interviewer, which may further compound this problem due to, for example, the necessity of travelling to and back from the interview location. Finding a mutually convenient time for researchers and participants to meet is a further challenge.

These issues may be alleviated through the use of telephone or online interviews, but, as Shuy (2002) explained, these lack "contextual naturalness" whereas "face-to-face interaction compels more small talk, politeness, routines, joking, nonverbal communication and asides" (p. 541). Face-to-face interviews may, thus, be more appropriate than other methods since these characteristics of face-to-face communication may mitigate any feelings of awkwardness and improve an otherwise tense atmosphere. However, this type of interview likely also places greater demands on researchers' interpersonal as well as oral questioning skills.

To alleviate the pressure on time and other resources, several individuals may be interviewed at the same time, but it would be difficult to achieve the same level of detail with each interviewee. Organising such group interviews is, logistically, significantly more challenging. Consequently, fewer interviews are likely (able) to be conducted unless the sample is small enough to be interviewed entirely with the available resources. Since this study's respondent pool was far larger than the available resources would permit to interview entirely, only a relatively small number of interviews was carried out. One key consequence of this is that insights from other participants not interviewed were

not captured and other possible themes in the data or correlations between the datasets were not uncovered.

Another pitfall, as highlighted by Sarantakos (2005), is the potential for the introduction of bias due to the presence of the interviewer who can, willingly or unwillingly, influence the interviewee. This may be all the more the case if the interviewer is perceived as a power figure who holds or is perceived to hold sway over certain aspects of the interviewee's personal, academic or professional life. For example, the interviewee may give responses which they believe are what the interviewer wants to hear, be it in an effort to garner favour or higher esteem from them, or to avoid possible enmity or other negative consequences in the future. This may represent a significant threat, in particular, to the confirmability of the findings since it cannot be guaranteed that interviewees' statements in such situations are not prejudiced (beyond a reasonable extent). This concern is discussed further within the Chinese context in Section 3.6.

Being aware of the teacher-student relationship in China as depicted in Section 2.9.1, it was, thus, decided to interview only those students who were no longer enrolled in any of the researcher's courses. This was a small sacrifice since this group represented the vast majority of the total number of participants. In addition, interviewees were encouraged at the outset of the interview to express themselves freely, and the researcher took great care during the interviews to avoid facial expressions or other gestures that may have been perceived by the interviewees as critical or disapproving.

A further issue is that, as discussed by Arksey & Knight (1999), interviews leave little possibility for anonymity *vis-à-vis* the researcher since the identity of the interviewee cannot easily be protected and since the interviewer may intrude into the interviewee's

personal environment (such as their home or office). These problems did not arise in this study, however. Respondents, as part of agreeing to an interview, provided their name and contact details, thereby consciously forfeiting anonymity. Furthermore, all interviews were conducted in a semi-public setting removed from the interviewees' personal sphere.

An additional characteristic typical of interviews that should be borne in mind is that the elicited information is based on recollections of past events. It is highly likely that the majority or perhaps all the questionnaire responses as well as those provided in the interviews were based on recollections of past events which may not be entirely accurate. While this is certainly a feature of any study which requires respondents to recall past events and how those events impacted them, it is important to acknowledge that inaccurate reports can call into question the validity or trustworthiness of the data collected (Brinkmann, 2013). Thomsen & Brinkmann (2009) made a number of suggestions and recommended interviewers to

> (1) allow time for recall and assure the interviewee that this is normal; (2) provide concrete cues […]; (3) use typical content categories of specific memories to derive cues […]; (4) ask for recent specific memories; (5) use relevant extended time line and landmark events as contextual cues […]; [and] (6) ask the interviewee for a free and detailed narrative of the specific memory (p. 303).

All of the above techniques were employed during the interviews conducted within the present study. To further stimulate participants' recall of their previous responses, an offer to provide a copy of their questionnaire electronically was made at the time when the date and time of interview were confirmed (in each case, approximately one week prior). In addition, a physical copy of the questionnaire was furnished at the interview. Before each interview question that pertained specifically to a question in the questionnaire, interviewees' responses were pointed to and briefly summarised; a few

additional moments were given to maximise the recall process. In order to evoke specifically the psychological impact of the feedback they had received, interviewees were invited to bring with them examples of marked written work.

3.12.3 Descriptive Metadata

The questionnaire (see Appendix 3) served as the entry point to exploring students' views on feedback. All questions in the interview schedule (see Appendix 4) were derived from students' responses to the items in the questionnaire. While also allowing additional themes to emerge, the central purpose of the interviews was to explore those already uncovered through the questionnaire more deeply.

As shown in Table 3.2, a high number of respondents provided answers to the open-response items in the questionnaire. Despite far fewer having provided answers to the eighth item, the total number of words compared to the sixth and seventh items was 29.4% and 63.5% higher, respectively. In addition, whereas the average number of words per respondent to the sixth and seventh items was approximately 16 and 13, respectively, that to the eighth item was 30 – 87.5% and 130.8% more, respectively. A possible reason for this is that the majority of respondents were also or more interested in commenting on other aspects of feedback not covered by the other items in the questionnaire. What seems clear is that they, generally, felt interested by the topic and that it was relevant to them, hence their motivation to offer additional comments in the eighth item.

Only the first 176 participants received the version of the questionnaire which contained a request for an interview since later participants were all enrolled in the researcher's own courses (this avoided a methodological concern discussed in Section 3.6). Of these,

96 (54.5%) agreed to be interviewed. In addition, 19 respondents requested a summary of the results. These numbers, again, indicate a keen interest on the subject of feedback. This is corroborated by some of the comments included in the e-mails received from the participants. One wrote: “I think your research about the feedback students have received from their Chinese teachers is interesting and meaningful” (permission to quote given). Several others included comments to this effect.

The questionnaire contained five closed items (no. 1-5) and three open-response items (no. 6-8) whose chief objective was to help answer the study’s research questions. These items and their corresponding research questions as well as response rates are summarised in Table 3.2.

Item no.	Question	Corresponding Research Question	Response Rate (*n* = 232)
1	How effective are the following types of feedback for helping you to improve your overall learning?	How helpful do the participants find each type of feedback they receive?	100%
2	How often do you receive each type of feedback on your assessments?	How frequently do they believe they receive each type of feedback?	100%
3	What is the overall effect of the feedback you receive on the following aspects?	How do they perceive the feedback they receive from their teachers to affect their academic performance, studying habits, motivation and self-confidence?	100%
4	When do you actually usually receive feedback on your assessments?	At which point(s) of the assessment process do they claim to normally receive feedback?	100%
5	When do you prefer to receive feedback on your assessments?	At which point(s) of the assessment process do they prefer to receive feedback?	100%
6	In my opinion, the most important purpose of feedback is ...	Which function(s) do they expect feedback to fulfil?	99.6%

7	Feedback is most helpful for me when …	How do they characterise effective feedback?	98.3%
8	Additional comments	Any	69%

Table 3.2 Questionnaire items with related research questions & response rates

Some additional explication of the three open-response questions is necessary in order to fully understand their purpose. As Chapter 2 has demonstrated, the functions of feedback, as well as of formative assessment and Assessment for Learning more generally, have been and continue to be central questions within the literature on educational assessment. The purpose of the sixth item was to help investigate participants' attitudes in this regard as well as to seek convergences and divergences between participants' and scholars' views.

The key purpose of the seventh item was to investigate what students considered to be "good" feedback, that is to say, that which was most beneficial for them. This item bears a close resemblance to the sixth item and, indeed, it was created, in part, as an adjunct thereto. In cases where respondents felt that the most important purpose of feedback coincided with the attribute(s) that render it most helpful for them (as some did), then their answer to the seventh item served to explore that to the sixth item more deeply. If this was not the case, the seventh item allowed them to expand on alternative aspects of feedback important to them which were not covered by the sixth item. In either scenario, the seventh item provided a greater opportunity for themes to emerge naturally. The seventh item used the word "helpful" since it is not too specific to lead or constrict answers but still specific enough to convey to participants that under investigation were those conditions which specifically lead to positive outcomes.

The intention of the eighth item was to allow students to express views or provide details on related issues not captured by the other questions. Responses to this item can be considered as particularly meaningful or powerful since: first, there were no parameters that guided respondents to answer the item in a particular way or to include particular content (although influence from previous items cannot be ruled out); and, second, there was very little if any obligation to answer it in the first place. Therefore, it is highly likely that participants' answers to this question represent that which was most important to them.

3.13 Conclusion

This chapter has engaged with various methodological aspects associated with the study, from the more theoretical to the more practical. A justification has been provided for the utilisation of two methods and for the emphasis on qualitative research and data which will become more apparent in the following chapter. Further, the chapter has contended with fundamental challenges faced by all researchers, including bias, the interpretation of reality and construction of knowledge, reliability and validity. Problems regarding conducting research specifically in the Chinese context have been considered as well. Additional more concrete elements of the study, including the pilot study, sample population and sampling methods, data analysis and coding strategies, ethical considerations, and the research instruments have also been discussed. All these aspects had the purpose of building the theoretical as well as practical framework on which to lay, interpret and present the evidence collected through the study. These findings are discussed in the following chapter.

Chapter 4

Findings and Discussion

4.1 Introduction

This chapter reports on and discusses the study's findings. The opening section investigates a possible link, suggested by a number of participants, between academic discipline and feedback frequency. The subsequent section presents and discusses the data collected through these two instruments. A number of themes identified through analysis of the data is laid out and examined in individual subsections. Additional observations, cross-references and links with the literature are made throughout. An overarching theme is discussed in the proceeding section. The final section ends the chapter by answering the study's research questions, recapitulating the primary findings and drawing some conclusions.

4.2 Link Between Field of Study and Feedback Type Frequency?

Before presenting and discussing themes identified in the data collected through the study, a potential relationship between students' main subjects of study and the frequency of different types of feedback they believe to receive will be explored. Such a relationship had not been identified in the initial literature search. However, as mentioned in Section 3.8, several students who piloted the interview schedule recommended taking into consideration their academic discipline.

Two further responses to the eighth item (which allowed respondents to offer additional comments) in the questionnaire likewise contained this suggestion. QU186's reasoning was that "students of different majors receive different assessments," setting as a

contrasting example English subject students and those studying science subjects. According to her, the former type of students received more annotations in their assignments than the latter. Similarly, QU220 explained that “different majors in [HI] may have different orientations in the form of feedback.” She distinguished between the social and natural sciences, and felt that students in the first category likely received more comments while those in the second likely received more grades.

Helping to substantiate this view retrospectively were explanations from two interviewees for the high frequency of all types of feedback they perceived to receive. QIU30, a social science student, stated that her classes were much smaller than natural science classes, and felt that her teachers would, thus, have “more time and energy” to provide feedback. QIU63, who reported often receiving a number of different types of written feedback, provided a similar explanation.

In light of these comments collected through the questionnaire pilot and main studies, it was decided to add a question to the final version of the interview schedule which asked interviewees to state their main subject of study. Fortuitously, precisely one half of the interviewees studied natural science subjects while the other half studied social science subjects so that a direct comparison is possible. In addition to testing whether this correlation actually existed, an additional motivation was that, if yes, it might have an important bearing on how the data should be interpreted or on how tutor feedback should be investigated in the future. Although speaking more broadly of formative assessment, Hodgen & Marshall (2005) made a similar realisation:

> If there were substantial differences between subjects then this would have considerable implications for any attempt at generic discussions of formative

> assessment. Indeed, if the differences were too great we might have to question the extent to which formative assessment was a meaningful concept (p. 156).

Figure 4.1 shows a comparison of natural and social science students' beliefs regarding the frequency of receipt of marks or grades. Roughly the same numbers believed they "always" or "never" received this type of feedback. While more natural science students felt they "often" received it, more social science students perceived to "sometimes" receive it. Since the comparison is only between two sets of 15 students, the explanatory power is minimal. However, as far as this group of students is concerned, the idea that students taking natural science courses receive feedback in the form of marks or grades more frequently than their counterparts taking social science courses is, generally, not true.

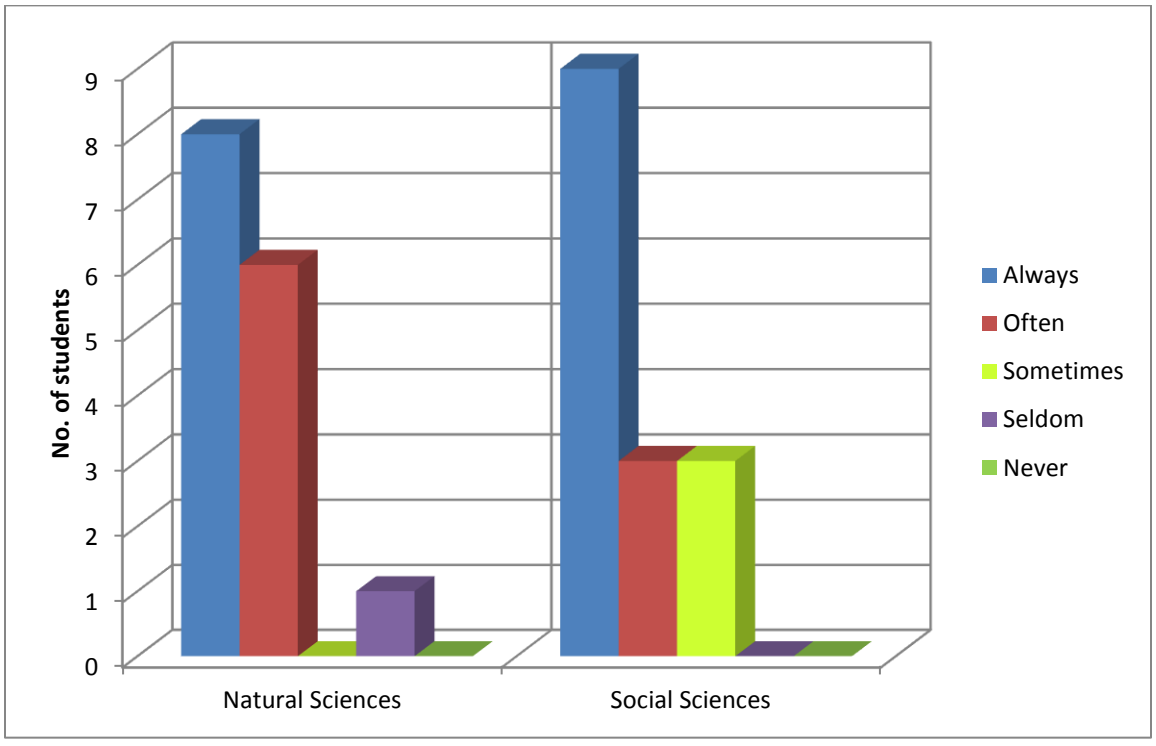


Figure 4.1 Perceived frequency of receipt of marks/grades

Figure 4.2 shows a comparison of natural and social science students' beliefs regarding the frequency of comments, i.e. all other types of feedback. The data shows that natural and social science students felt that they received feedback in the form of comments with roughly the same frequency. While fewer natural science students believed they "often" or "seldom" received comments, fewer social science students perceived themselves to "sometimes" or "never" receive them. Thus, the idea that natural science students receive comments less frequently than their counterparts from the social sciences is, generally, also not true.

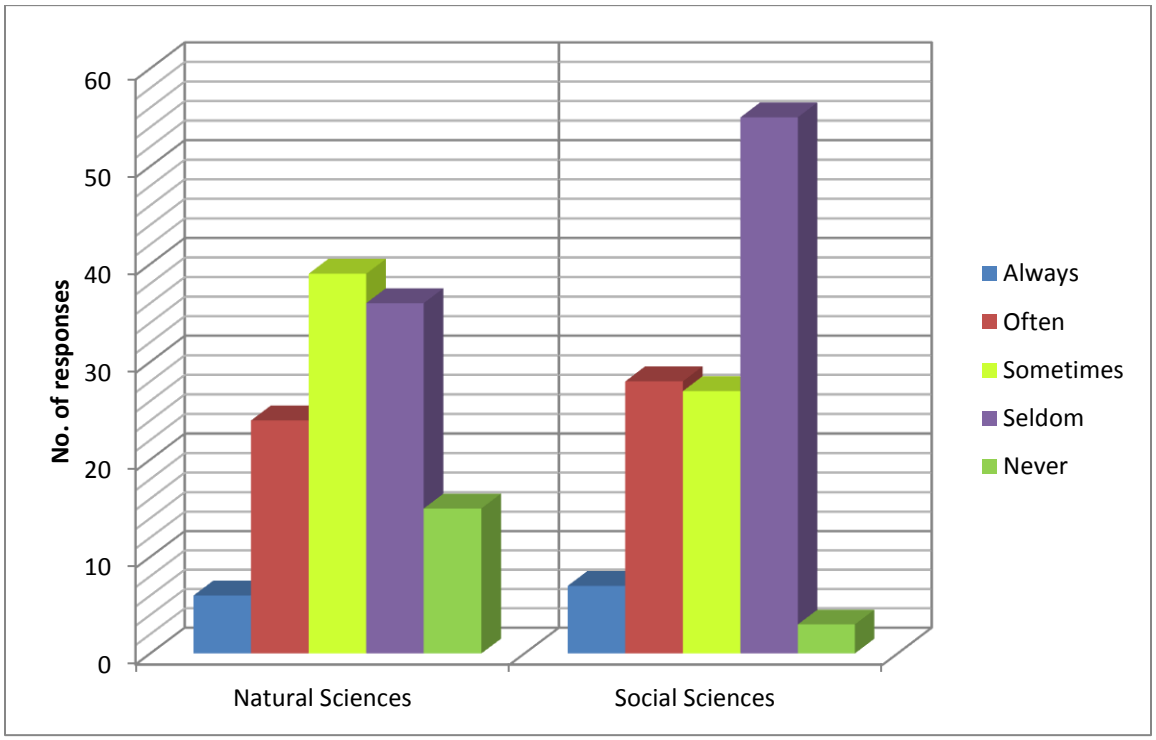


Figure 4.2 Perceived frequency of receipt of comments

These findings may be surprising considering the accounts of students cited previously. However, it emanated from some of the interviews that natural science students were required to take a number of mandatory social science classes and *vice versa*. In addition, it should be borne in mind that students' reports of the feedback they received

encompassed all of their courses rather than only those related to their main subjects of study. In consideration of these two facts, it stands to reason that a contrastive examination, such as the one undertaken here, would likely not produce major differences.

In spite of this conclusion, it would be erroneous to assume that there is no relationship as such between students' main subject area and the frequency with which they (believe to) receive different types of feedback, be it generally, more specifically at the host institution or even within the interview sample. More variables would need to be taken into consideration for the existence of a correlation to be confirmed or refuted with a higher degree of confidence. An example of a possible mediating variable is students' year of study: students in different years of study may (tend to) take different numbers of courses related to their main subject area. Thus, for instance, a group of second-year social science students may (perceive to) receive comments less frequently than a group of their peers who are in their fourth year of study.

Subject-specific investigations into formative assessment are in the minority. Black & Wiliam (1998a) highlighted this shortcoming in several of the studies included in their paper. Other authors (e.g., Bloxham & Campbell, 2010; Bennett, 2011) also averred that not enough attention is given to the subject domain. Writing on self-assessment in religious education, Brooks & Fancourt (2012) asserted that "much of the literature is generic, displaying limited attention to the particular subjects involved, even though the curriculum in most schools, further and higher education is subject-based" (p. 124). The authors used English and mathematics to exemplify differences in various pedagogical practices between these two disciplines.

Hodgen & Marshall (2005) systematically compared formative assessment practices in these two subject areas which "are often thought of as contrasting disciplines, typifying the arts/sciences split in education" (p. 154). They concluded that what made the English and mathematics lessons they observed in their study formative was "largely the same" and that the "pedagogic principles of formative assessment appear to be generic" (p. 172). Despite this, they refrained from attempting a comparative analysis since they found themselves "often talking at cross-purposes" (p. 155), for instance by using concepts from one of the disciplines which was incompatible with the other or by using the same terms but in reference to different phenomena. Alternatively, they recommended "adapt[ing] the generic principles of formative assessment […] to the more specific demands of each school subject" (p. 172). A similar recommendation could be made in relation to investigations into frequency and other aspects of feedback in higher education. Section 4.3.2 further discusses feedback frequency.

4.3 Themes

4.3.1 Effectiveness of Different Types of Feedback

The first item investigated through the questionnaire was students' views on a range of feedback types in terms of their effectiveness in helping them to improve their learning. Figure 4.3 provides a summary of the total number of responses for each type of feedback (in the case of this and other figures, descriptors on the horizontal axis have been abbreviated to fit the chart).

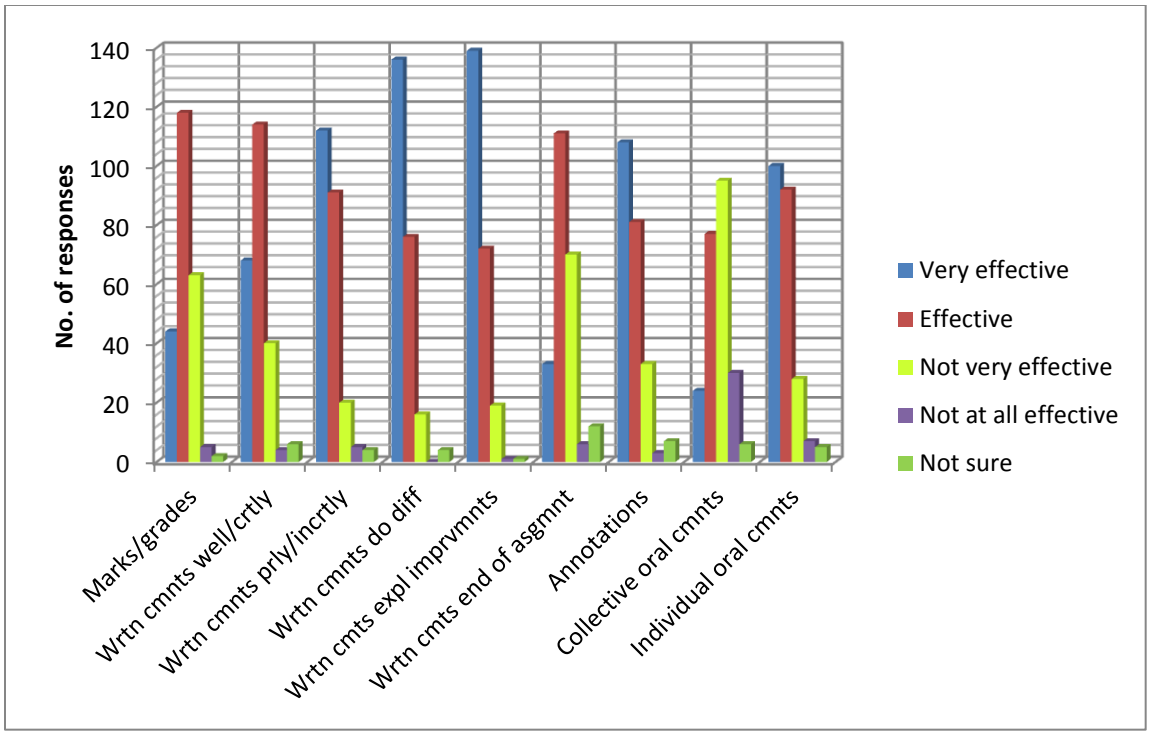


Figure 4.3 Differentiated effectiveness of different types of feedback

The results show that the respondents held a strong preference for written feedback. The overwhelming majority considered written comments that tell them what they have done well or correctly (78.5%), those that tell them what they have done poorly or incorrectly (87.5%), those that tell them what they need to do differently in the future (91.4%), those that explain how to make improvements (91%), and annotations in the body or margin of their written assignments (81.5%) to be either “effective” or “very effective;” most also provided these ratings in the case of comments written together at the end of written assignments (62.1%). In addition and as the chart illustrates, very few students deemed any type of written feedback to be “not at all effective.” A strong majority also held oral comments given to them directly about their work during or after lessons to be either “effective” or “very effective” (82.8%).

As the chart shows, only one type of feedback – oral comments given collectively to groups or the whole class during lessons – was rated most frequently as "not very effective." This type of feedback also received the highest number of responses labelling it as "not at all effective." Two types of feedback – marks or grades and comments written together at the end of written assignments – received mixed reviews. In the case of the former type of feedback, 19% of respondents felt it to be "very effective" while 27.2% rated it as "not very effective." In the case of the latter type of feedback, a high proportion of respondents felt it to be "not very effective" (30.2%) while, contrastingly, a smaller but still notable number considered it to be "very effective" (14.2%).

With respect to written comments in general, respondents expressed a preference for this feedback type over marks or grades, which reflects conclusions drawn in major reviews, such as those from Crooks (1988) and Black & Wiliam (1998a). Students' stronger preference for written comments on how to make improvements than for those that tell them what they have done well or correctly or those that tell them what they have done poorly or incorrectly also coincides with other researchers' findings that feedback on how to improve an answer rather than simply whether it is right or wrong is more effective (Shute, 2008).

Sully de Luque & Sommer (2000) claimed that students from Confucian-heritage countries – such as China – prefer indirect, implicit and group-based feedback. Conversely, in terms of the effectiveness of oral comments, the participants reported a preference for such feedback to be given individually rather than to groups or the whole class. Students from Confucian-heritage countries may prefer direct, explicit and individual feedback but only if it is given in private. This would likely apply particularly in

the case of critical comments since, if these relate to a specific student and are given in front of the entire cohort, that student may lose face (see Section 3.6).

By assigning each of the effectiveness categories a numerical value, it is possible to compare the effectiveness of the different types of feedback included in the questionnaire in a way which more clearly illuminates similarities and dissimilarities. The "very effective" figures were multiplied by a value of 2; the "effective" figures by a value of 1; the "not very effective" figures by a value of -1; and the "not at all effective" figures by a value of -2. The figures in the "not sure" category have been omitted since their numerical value is unknown or not quantifiable. The results are shown in Figure 4.4. The vertical axis has been increased to the maximum possible score of 464 in order to demonstrate the absolute as well as relative effectiveness of each of the feedback types.

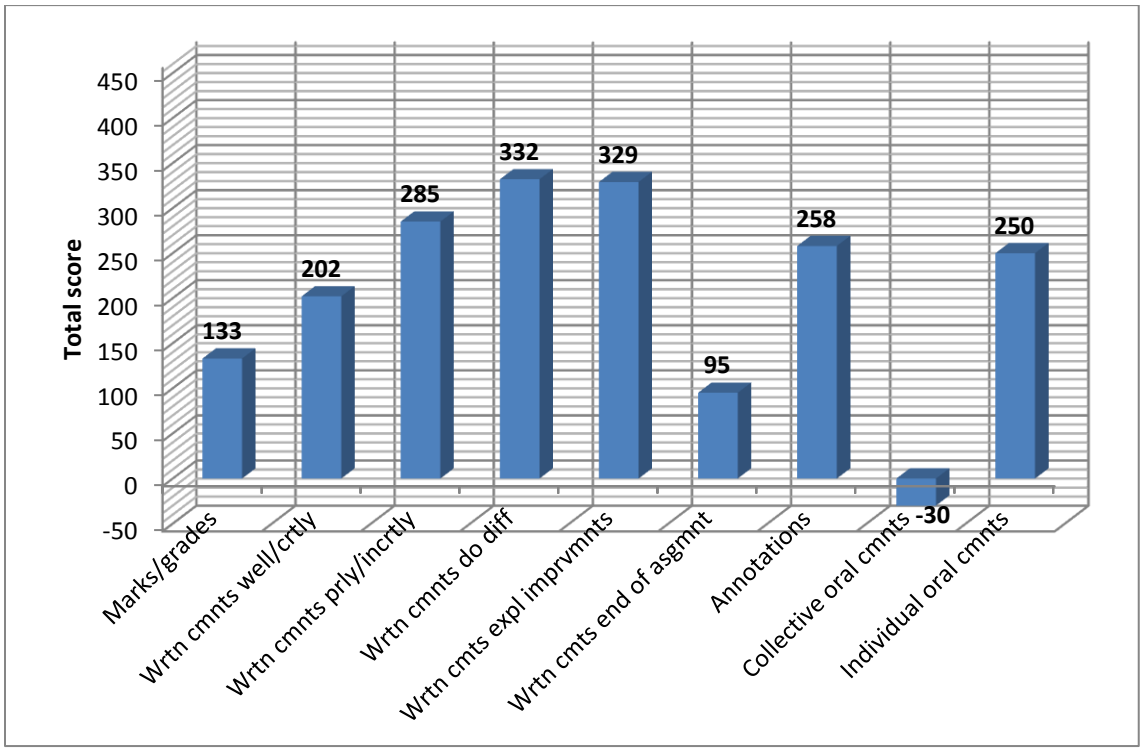


Figure 4.4 Scored effectiveness of different types of feedback

While the values are not meaningful in themselves, they do reinforce the finding that students were generally satisfied with the types of feedback they received in terms of effectiveness. They also show the form of feedback rated as most helpful to be written comments that explain what needs to be done differently in the future. Almost on par with this type of feedback is written comments that explain how to make improvements. Oral comments given collectively to groups or the whole class during lessons received the lowest score by far. “Marks/grades” received a relatively low score but still higher than comments written together at the end of written assignments.

Since marks and grades are, by definition, summative (Lipnevich & Smith, 2009), it was decided to, where relevant, add a question to the interview schedule which asked how they might improve learning. Five of the eight interviewees questioned in this connection explained that they could be used to either confirm that the currently employed learning practices (such as time spent studying) were adequate and, thus, did not need to be altered, or not, in which case changes to the learning regime needed to be made. Two other students reported the same practice but additionally explained that they referred to the class average to determine whether their result was satisfactory. IU152 stated that receiving a positive result on an assessment made him feel “happy” and “confident” while a disappointing result motivated him to “try harder the next time.”

These would appear to be examples of summative assessment being used for formative purposes or producing formative outcomes, and are strongly reminiscent of Tara’s (2002) position that “marks have a place even in formative assessment” (p. 507). This also coincides with Bloom’s (1969) earlier stance (see Section 2.4) that assessments are neither necessarily formative nor necessarily summative, even though formative evaluation is *better* supported if it is detached from the grading process.

A discreet categorisation of assessments into *de facto* formative or summative ones may even be fallacious. This point has been a central matter of discussion in the literature. Black (1998), for example, called attention to the fact that

> Some have laid stress on the differences between the formative and summative purposes, and have argued that the assessment instruments and procedures needed for the one are so different from those for the other that neither can flourish without clear separation. On the other side, it can be argued that the two functions are two ends of the same spectrum and that there is no sharp difference, and that if the two functions are separated, then teachers' assessment work will be devalued (p. 34).

However, even the notion of purpose may become obscured depending on, for example, the level at which it is examined. Newton (2007) argued precisely this point and proposed three broad interpretations of assessment purposes: to derive standards-referenced judgements, to support selection decisions and to produce certain (intended) impacts on the students, such as the safeguarding of their motivation (p. 150). As shown in Section 4.3.3, many participants felt the latter element to be a particularly important characteristic of effective feedback.

In contrast to Black (1998), Newton (2007) appeared not only to place the terms "formative" and "summative" at opposite ends of a spectrum, but on different spectra (that is, in different categories) altogether:

> the term 'summative' can only meaningfully characterize a type of assessment judgement (i.e. it operates at the judgement level of discourse), while the term 'formative' can only meaningfully characterize a type of use to which assessment judgements are put (i.e. it operates at the decision level of discourse). The terms belong to qualitatively different categories; to attempt to identify characteristics that distinguish them – within a single category – is to make a category error (p. 156).

Thus, as he went on to explain, there are, in fact, no summative purposes or formative judgements. He admitted earlier, however, that his three categories also needed to be clearly distinguished, and that, where this was not done, “their distinct implications for assessment design may become obscured” (*ibid*).

Hence, the labelling of assessments as either formative or summative is problematic and supports this thesis’ proposition that the decisive factor be on the outcome. This may also be more amenable to the conceptualisation of formative assessment across different national or cultural realms, such as China and other Confucian-heritage countries, where notably different assessment histories and practices exist. Furthermore, feedback produced through an assessment instrument that led not only to changes or to decisions of some kind but also specifically to positive ones – such as those identified in the eight students' responses above –, ought to be promoted from formative feedback to feedback for learning.

Thirteen interviewees who had answered that they found marks or grades to be “not very effective” or “not at all effective” were also questioned on this point. More than three quarters explained that they lacked specificity and that they could not inform them as to in which areas they had performed well or poorly. QIU15, for example, stated:

> You know that you are wrong or right, but you don’t know why.

Thus, interviewees who had answered that marks or grades were an effective form of feedback were asked how they could know in which areas(s) specifically they needed to improve. All eight responded that the instructor usually permitted them to view their marked assessments so that they could see which questions they had answered

incorrectly or in which section(s) they had lost points. However, it is unclear how this could help them understand *why* they had made a particular error and how they could avoid repeating it in the future. This point was not pursued further in the interest of proceeding with the interview.

The issue of lacking specificity was also the reason given by the majority of students who were asked about their having rated oral comments given collectively to groups or the whole class during lessons as "not very effective" or "not at all effective." For instance, IU19 pointed out that not only were oral comments difficult to retain or remember accurately, but they were also too broad to be of any value to individual students. Carless (2006) noted that lacking specificity is one of the key criticisms directed by students in higher education specifically at the feedback they receive. This is also consistent with Shute's (2008) comment that feedback which lacks specificity may cause pupils to consider it to be useless or to feel frustrated.

IU19 and other students explained that this was also why they preferred written comments since these were individual and could be referred to any number of times again in the future. Hence, it would be expected that students found annotated comments to be more helpful than those written together at the end of an assignment since these are more explicit and exemplify a particular error or suggestion for improvement. Indeed, Figure 4.4 shows that the former feedback type has a value almost three times that of the latter.

Written comments also have psychological benefits. IU108, for example, explained:

> Annotations make me feel important and that the teacher cares about what I did.

In contrast to IU19, IU47 felt that oral comments given collectively to groups or the whole class during lessons were helpful in the case of a comment being made about a widespread problem on an assessment that did not apply to him; in this manner, he could ascertain whether he was ahead of most of his peers.

Those 8 students who had answered that they found all types of feedback to be "effective" or "very effective" were also questioned again. A diversity of reasons was given. For example, IP57 explained that she frequently received no feedback at all and that:

> Compared to no comments, any kind of assessment is good.

QIU18 stated that it was because of receiving significantly less feedback since entering university and that feedback was, therefore, something to be cherished, regardless of which form it came in. IU81 explained that any type of personal feedback gave her the feeling of being noticed, which acted to motivate her.

These responses indicate that the students were accustomed to receiving feedback more frequently or more individualised comments prior to entering university and, now, have a greater appreciation for feedback owing to being given far less or fewer personalised comments. They also suggest that the efficacy of feedback – when it takes the form of (any type of) comments – can, in some cases, rest on the frequency with which they are given rather than on which specific form such feedback is provided. In other words, the more comments, the better.

4.3.2 Frequency of Different Types of Feedback

The second item in the questionnaire investigated students' perceptions of the frequency of different feedback types. Figure 4.5 provides a summary of the total number of responses for each type of feedback.

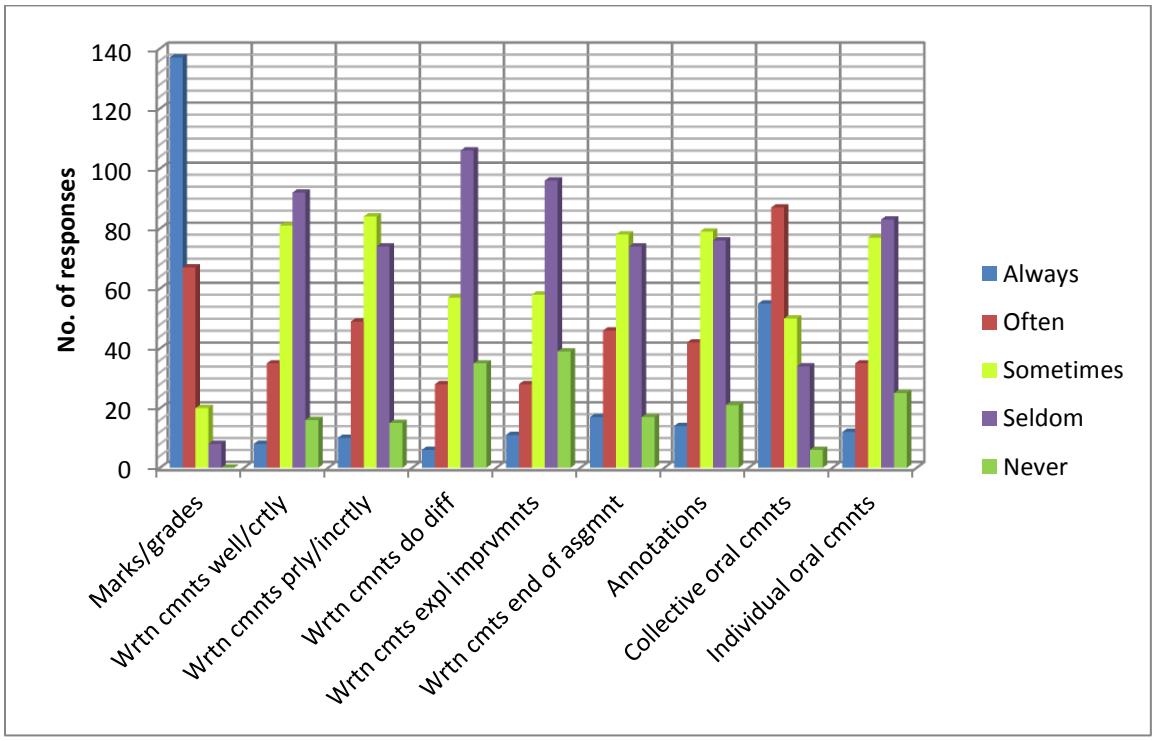


Figure 4.5 Differentiated perceived frequency of different types of feedback

By far, the most common feedback type, according to the participants, is "marks/grades," with the resounding majority reporting that they believed they either received it "often" or "always" (87.9%). No students answered that they felt they "never" received this type of feedback. Similarly, the majority believed that they received oral comments given collectively to groups or the whole class during lessons either "often" or "always" (61.2%), though to a lesser extent than marks or grades.

The majority of respondents believed they received all other types of feedback either "sometimes" or "seldom." A particularly high proportion felt that they "never" received written comments that tell them what they need to do differently in the future (15.1%) nor those that explain how to make improvements (16.8%). As reported in the previous section, these types of feedback were also assessed by the large majority of students to be either "effective" or "very effective." This phenomenon reflects the situation in higher education in the U.K. According to Nicol (2010), whereas

> written feedback was part of a larger coordinated system of teacher–student communication, […] in most institutions, due to the growth in student numbers, written comments have become detached from this supportive context (p. 501).

As in the case of the previous theme, each frequency category can be assigned a numerical value, which enables a clearer comparison of the perceived frequency of the different types of feedback included in the questionnaire. The "always" figures were multiplied by a value of 4; the "often" figures by a value of 3; the "sometimes" figures by a value of 2; the "seldom" figures by a value of 1; and the figures in the "never" by a value of -1. The results are shown in Figure 4.6. The vertical axis has been increased to the maximum possible score of 928 in order to demonstrate the absolute as well as relative frequency of each of the feedback types.

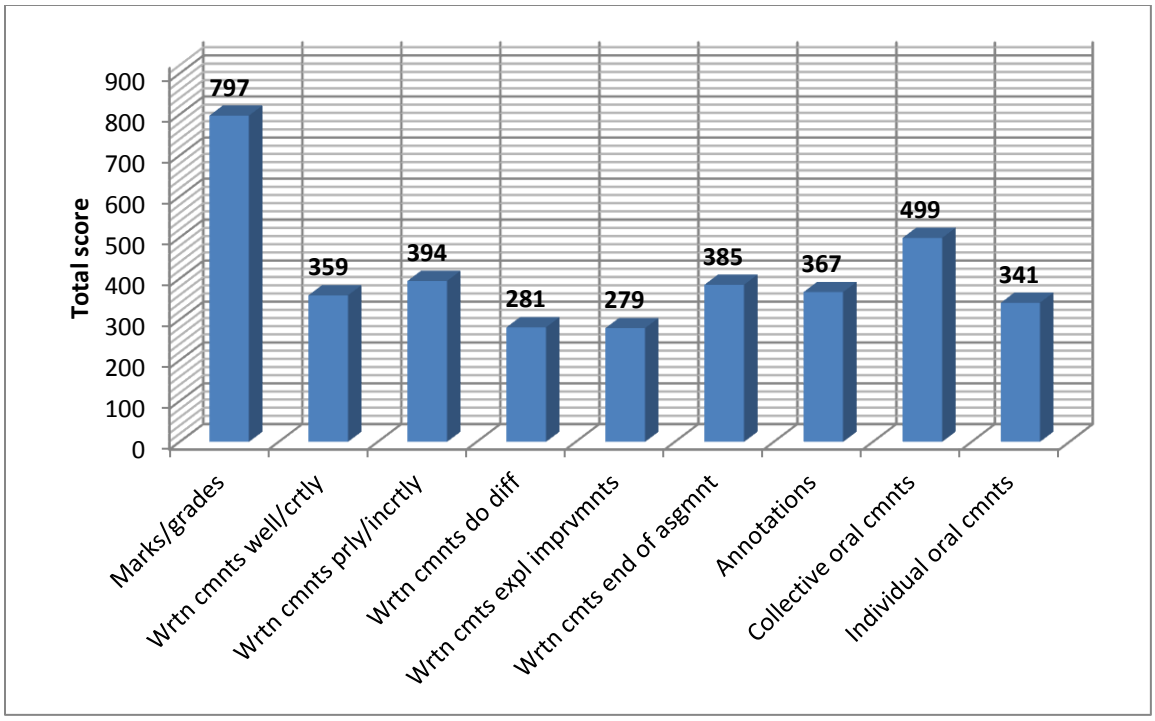


Figure 4.6 Scored perceived frequency of different types of feedback

Again, while the values themselves are not meaningful, they do clearly show that the form of feedback students supposed they received most often is “marks/grades;” no other feedback type is comparable. However, oral comments given collectively to groups or the whole class during lessons also stand out as being regarded as a frequent form of feedback. The types of feedback respondents stated they received most infrequently are written comments that explain what needs to be done differently in the future and those that explain how to make improvements.

The fact that students in China or any other part of the world receive feedback in the form of marks or grades is hardly surprising since, logically, a summative assessment has to be made at some stage during or at the end of an academic course. The question is one of relative frequency: why are marks or grades and oral comments given collectively to groups or the whole class during lessons so much more frequent than all

the other types of feedback? IP138 answered that the reason was one of “convenience” for the teachers who, for example in a mathematics class, typically faced 200-300 students; they were unable to provide individual feedback simply for practical reasons. IU156 explained that, for the same reason, teachers frequently spoke in front of the class about “collective problems” rather than addressing each student about their individual needs. However, even these comments were, according to QIU15, based on information reported to them by teaching assistants who were employed primarily to mark students’ homework.

All other interviewees questioned on this subject answered that classes usually contained too many students, that the teachers did not have enough time or that they were too busy to provide students with individual feedback that went beyond marks or grades. QIU6 noted that, due to large class sizes:

> It’s really hard for the teacher to grade each assignment, let alone for him to give a comment.

She added, however, that, even in classes related to her major (i.e. main subject), which may contain as few as 20 students, the situation was similar; she did not know why. This calls into question the infrequent provision of individual oral feedback and written comments as being necessarily associated with the logistical challenges of large class sizes. Several other students stated that teachers were willing to provide individual feedback if they were sought, for example, during their office hours but that it was often difficult to make an appointment because of high demand.

In contrast to these students, two interviewees had answered that they “often” or even “always” received other types of individual feedback. IU108 described feeling

"surrounded by all kinds of feedback every day." He appeared to have a broader understanding of feedback which included virtually all communication from his teachers, including "body language" and "facial expressions." IU81 explained that she actively requested feedback, for example, by addressing a question to the teacher at the end of an assignment. She stated:

> What is most important is not the submission of my homework […]; what really counts is the feedback from the teachers so that I can improve myself.

The elucidations from these two students demonstrate the high potential for subjectivity in the interpretation of both the nature and the frequency of received feedback. While this is relevant in readers' estimation of the findings of this study, it is also important in that of others. Learners' interpretation of the research subject and of aspects related thereto may go well beyond that of the researcher's.

4.3.3 Psychological Effects of Feedback

This theme represents a very substantial part of the study's findings and discussion due, in great part, to its multifaceted nature. It can be considered as an umbrella theme since it encompasses a number of individual affective themes, some of which are interrelated and overlap. A further reason for its prominence in this chapter is the fact that, in addition to having been explored in the study by design, it surfaced as an important theme in all three open-response questions in the questionnaire.

This mirrors a study from Xiao & Carless (2013) which investigated 29 purposively selected high (i.e. senior secondary) school students' experiences of English language assessment through drawings and interviews. Whereas the authors found that students' negative perceptions of assessment clearly outweighed positive ones, their positive

perceptions mainly fell into two categories, both of which are related to the affective domain: “satisfaction with attainment, such as high marks and praise, and feelings of being supported or encouraged by teachers and peers” (p. 16). Furthermore, two of the three key functions Shute (2008) ascribed to formative feedback, as discussed in Section 2.8, relate to the affective domain.

In the third item in the questionnaire, participants were asked to state the overall effect of the feedback they received on four different aspects. Figure 4.7 provides a summary of the total numbers of responses for each aspect.

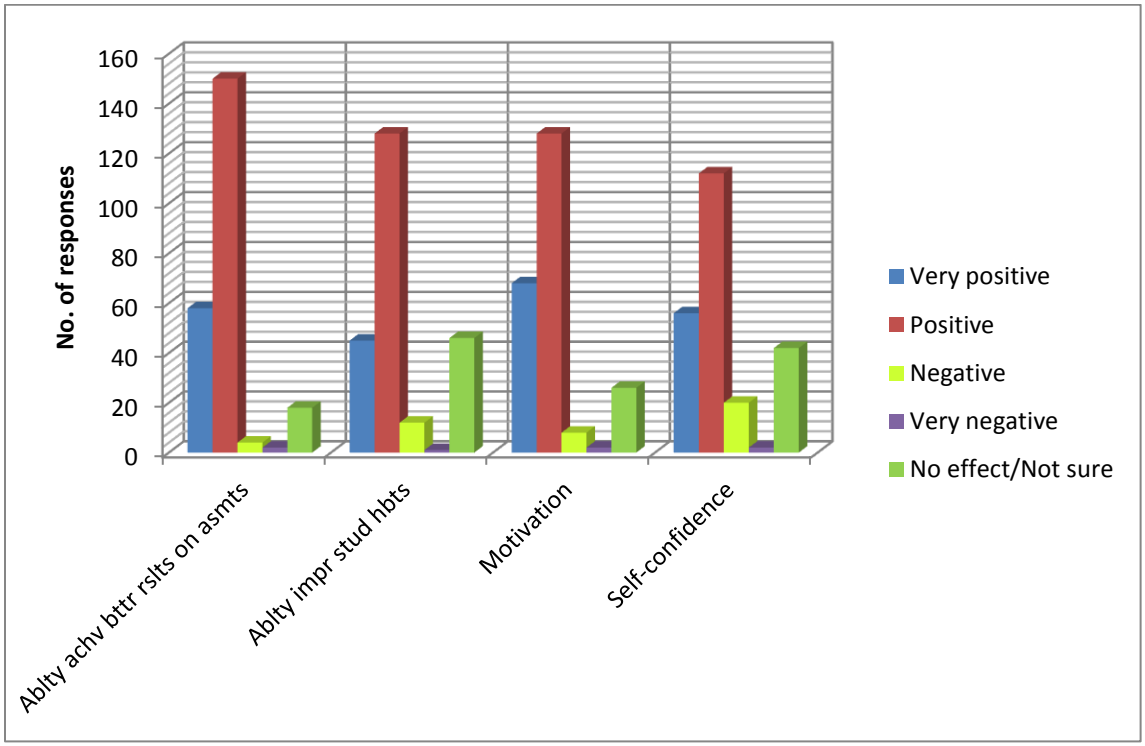


Figure 4.7 Effect of feedback on four aspects

Clearly, a considerable majority of students felt that the feedback they received influenced all four aspects positively; a relatively large proportion also reported the effect to be “very positive” in each case. In none of the four aspects did many participants

evaluate the effect to be “negative” and even fewer “very negative.” However, there also appeared to be some uncertainty concerning, in particular, the effect on their ability to improve their studying habits and on their self-confidence, with a sizeable minority selecting “no effect/not sure.” The former finding reflects the low frequency with which respondents felt to receive written comments that tell them what they need to do differently in the future and those that explain how to make improvements since these two forms of feedback are most closely linked to the improvement or modification of studying habits. This aspect emerged as a theme in its own right and is discussed further in Section 4.3.5.

A total of 33 (14.4%) responses to the sixth item (which asked respondents to state what they considered to be the most important purpose of feedback) in the questionnaire can be attributed to the psychological theme. There is a number of more specific aspects within this theme that emerged during the analysis of students’ responses to the sixth item. The most numerous of these include: approval or attention from the teacher; confidence; interest; and encouragement or motivation. Connecting all these aspects is that they can lead to improved learning outcomes, albeit indirectly. The following selection of responses encapsulates students’ desire for the feedback they receive to fulfil particular psychological requirements:

Approval/attention from the teacher:

> to give the students a feeling that the teacher cares what they’ve been doing in the task, and that the task is significant to their self-improvement. More of an encouragement in this sense, equally important with the actual assessment teachers have made (QU25).
>
> to let the student feel noticed and paid attention to by the teacher (QU31).
>
> to make students feel that they are cared by their teachers […] (QU116).

Confidence:

> to know the field that I am weak in and show me how to study deep in it, and also to improve my confidence in study (QU13).
>
> to help me do better in the future assignment, and also to give me advice and confidence (QU126).
>
> to give students confidence to achieve their rest assessment (QU214).

Encouragement/motivation:

> to motivate students to achieve better performance in the future (QU53).
>
> to encourage students and help them make improvements, which may be a great incentive for students to be more confident and highly-motivated (QU69).
>
> to motivate students to improve themselves and help them do better (QU89).

Interest:

> to make students interested in this course (QU7).
>
> to help a student have deeper understanding of a concept and have more interest in thinking and solving a problem (QIU15).
>
> to inspire my interest in the subject and point out certain aspects of doing the assignment that I left out (QU109).

The psychological theme also featured strongly in students' answers to the seventh item (which asked respondents to describe when feedback was most helpful for them) in the questionnaire, with 22 (9.7%) responses exhibiting related aspects. The most numerous of these include: confidence; depression or frustration; and encouragement or motivation. The first and third of these aspects overlap with the responses to the sixth item while the second one is new. Some examples of each of these aspects include:

Confidence:

> I lose confidence in the class (QU159).
>
> I lose my passion or self-confidence (QU193).

> I was not so confident and began to doubt myself (QU204).

Depression/frustration:

> students are in a bad mood or feel frustrated (QU37).
>
> I get stuck in one tough situation and feel terribly depressed (QU115).
>
> I feel low or I feel depressed. In [HI], maybe many or most of us have much pressure from study and competition, so we study hard. But sometimes I may feel unconfident and depressed, I need the helpful feedback from my teacher, sometimes which is a piece of encouragement or a right direction may help me much (QU220).

Encouragement/motivation:

> the feedback can arouse my interest in the subject and I can be motivated (QIU30).
>
> I feel confused about my ability on this course. For instance, I need the encouragement from instructors as 'Well done. you have a good and profound understanding. However, you may make more efforts on grammar and vocabulary' (QU155).
>
> my motivation of study is in lower grade, feedback is helpful for me because it makes my notice that I've done something (good/bad) and the teachers know about it (QU189).

Affective aspects manifested themselves as a recurrent theme amongst 17 (10.6%) responses from students to the eighth item (which allowed respondents to offer additional comments) in the questionnaire. In this case, all responses in this category revolved around either confidence, or encouragement or motivation, both of which also emerged in students' answers to the sixth and seventh items in the questionnaire.

Confidence:

> Since I found that professors here are less "individual based" than high school teachers, and I need more feedback to help me make improvements and build up confidence (QU102).
>
> Although I think it important to praise students in the feedback, helping students find their errors without hurting their confidence is of great significance as well (QU200).

> Make the feedback more direct and to the point, and don't worry that these caustic feedbacks will frustrate students' confidence because we students now need more setback to build stronger heart (QU232).

Encouragement/motivation:

> Feedback is necessary for students to keep on the right way in the ocean of knowledge and not lost themselves. Additionally, feedback also plays an important role in motivating students in spiritual level, which may urge them not to give up (QU28).
>
> In my opinion, the feedback had better focus on the right way instead of the right answer. And it will be helpful if the feedback includes some encouragement (QU176).
>
> I wish to have more comments about my own assessment from teachers, not just grades as I don't know my performance and way to improve just by scores. And in comments, maybe some praise words can motivate me much more than just recommendation (QU227).

The reappearance of these affective elements validates Black & Wiliam's (Assessment Reform Group, 1999) and the Assessment Reform Group's (2002) emphasis on considering the emotional impact of assessment on students. This finding also strongly suggests that affect is, at least in the views of this study's participants, a centrepiece of feedback for learning. In light of this, it is astounding that only 3% of the articles on assessment and feedback in higher education Evans (2013) included in her meta-review had the affective domain as their central theme. Dowden *et al*. (2013) agreed that "little attention has been brought to bear on students' emotional responses to feedback" (p. 349).

The role of affect has, however, been acknowledged by some scholars. For instance, Värlander (2008) maintained that "Students' emotions greatly influence the way in which they are able to receive and process feedback," and that the usefulness of such feedback may sometimes be overridden by their reactions' to it (p. 146). Likewise, Gamlem & Smith (2013) averred "How students interpret and deal with feedback is

important information to make formative assessment support learning, and involves both psychological states and dispositions" (p. 150). Nicol & Macfarlane-Dick (2006) also recognised the importance of feedback in playing a psychologically supportive role by including this element in their seven principles of good feedback practice. According to the authors, such feedback should encourage positive motivational beliefs and self-esteem which

> are more likely to be enhanced when a course has many low-stakes assessment tasks, with feedback geared to providing information about progress and achievement, rather than high-stakes summative assessment tasks where information is only about success or failure, or about how students compare with their peers (e.g. grades) (p. 212).

The reality in Chinese classrooms is, as portrayed in Sections 1.3 and 2.9.1, that the assessment landscape is dominated by high-stakes, summative testing. This is corroborated by the findings presented in Section 4.3.2. Therefore, according to Nicol & Macfarlane-Dick's (2006) logic, it would be expected that these affective elements would be a recurring theme in a study investigating Chinese students' experiences of and views towards the feedback they received.

This theme was further investigated through the interviews. The purpose of doing this was to help uncover exactly which characteristics of the feedback students received influenced the four factors that were included in the questionnaire so positively.

IU19 explained that her responses were based primarily on one course which took up a significant part of her weekly schedule. It was attended by approximately 10 students, and the teacher frequently met with each of them and provided high-quality feedback that was not overly critical and was encouraging, specific and focused on helping them

improve. IU26 stated that his responses were heavily dependent on the type of feedback, in particular marks or grades. He explained that, in the case of a poor result, his self-confidence might be damaged, but his motivation would increase due to high expectations as well as heightened pressure to achieve better results.

When explaining why his motivation and self-confidence were affected positively by the feedback he received, IU104 recalled an occasion during a class when the teacher addressed him and praised him for his improvements on his assignments. He exclaimed how surprised he was that the teacher even knew him considering that the class contained over 200 students. He added that, even though he felt that the class itself was useless for his future work, the feedback he received benefitted him greatly. He further explained that the impact of the praise he received was all the greater in light of the respect that teachers in China command.

IP167 also spoke of how teachers in China were generally held in high esteem but that this was often a hindrance to students' obtaining feedback since they lacked the confidence to approach them. This student candidly stated:

> Many are afraid of the teachers.

This supports Yang & Carless' (2013) statement that "Perceived unequal power relationship with teachers can cause students to lose confidence in obtaining teacher feedback" (p. 289). Similarly, Higgins (2000) argued that "the process of giving and receiving feedback needs to be understood within the context of a particular tutor-student relationship based on an unequal distribution of power where the tutor is accepted by students as an authority figure and expert" (para. 17). Thus, investigations

into the affective impact of tutor feedback on students, in particular in environments where a considerable imbalance in power between students and teachers is likely to exist, would be more valid and more illuminating if they offered a detailed depiction of the social or cultural context as this study has done.

Värlander (2008) explained that perceived power asymmetries in feedback situations are particularly pronounced in the case of tutor-student interaction "due to the dual role of the tutor as both assisting and passing judgement on the student" (p. 152). These may produce a number of negative emotional impacts on students, such as fear, anxiety, low self-esteem and diminished motivation. One suggestion the author made to help counteract these effects is to make greater use of peer-assessment which redirects power away from the tutor to the students. To the researcher's knowledge (from conversations with Chinese instructors), this form of assessment is utilised at the host institution, but the manner in, extent to and success with which this is done is not known. This practice is likely not a widespread phenomenon in China and may not be for some time owing to the deeply-rooted summative paradigm and teacher-led system of assessment that continue to prevail there (see Sections 1.3 and 2.9.1).

Värlander also stressed the importance of tutors showing greater empathy towards their students in order to reduce perceived power asymmetries. To achieve this, she recommended tutors engaging their students in discussions which prepare them for giving and receiving feedback. These activities give students the opportunity to voice their anxieties about and opinions on feedback as well as to better understand the role of feedback. To further help students in being able to understand the role of emotions in the provision and receipt of feedback, the author proposed tutors sharing their personal as well as professional experiences as well as case negative and positive case

examples of feedback with their students. According to the scholar, this may go some way in cultivating a positive emotional atmosphere that is more supportive of students' confidence and more amenable to interaction between them and their tutors. In principle, these meetings may work in any setting, but they would face greater challenges in those where class sizes are large and teaching resources are limited, such as in China, since they would require time to be set aside from the regular curriculum.

Ten students (4.3%) who had reported a "positive" or "very positive" effect of the feedback they received on all four of the aspects included in the third item of the questionnaire had also answered that they infrequently received comments. Those students in this group who were interviewed were asked how marks or grades affected them positively in these ways, but the majority replied that their answers were, in fact, based on the little feedback not in the form of marks or grades they *had* received. IU104 confessed that this was the case as did QIU8 who explained that any comments, which were not entirely negative, were beneficial to her.

IU156 also admitted that her answers were based on a small number of cases where she had received comments from her teachers at university. She produced a written assignment during the interview which she pointed out was the only one of its kind due to containing detailed comments. She felt inspired by the fact that the teacher had done this, demonstrating that they "care about" her work instead of simply writing "academic garbage," referring to what Carless (2006) termed "academic discourse" – the language in which teachers' comments are (often inaccessibly) couched. She reminisced about the feedback and individual attention she and her peers received from their teachers in high school.

This reflects Yang & Carless' (2013) remark that students' desire for timely and effective discourse about their improvement with their instructors is often left unfulfilled since "the university experience cannot reproduce the sustained support of the secondary school" (p. 286). Thus, individual incidences of direct contact such as IU156's may be all the more evocative and valuable. QIU8 went so far as to say that "what the feedback means is not so important," but rather more simply the fact that they received feedback which meant that "there is someone who is paying attention to you." She added that this gave her a "sense of importance" and the feeling that her work was "meaningful."

Similar reactions to feedback were collected though an investigation by Lipnevich & Smith (2009) into students' views towards different types and combinations of feedback at a university in the U.S.A. Participants were divided into six focus groups, each one containing eight or nine students and receiving one of the following types and combinations of feedback: tutor feedback with a grade; that without; computer-generated feedback with a grade; that without; only a grade; or no feedback of any kind. Each group also included an equal number of participants who did and did not receive praise.

The researchers reported that students in the second group expressed delight at the individual attention they received from the teacher, using quotations like "'I was, like, cool, he took time to help me better my essay!'" (pp. 355-356) and "'I never got so much feedback, it's so useful but nobody ever does it. I couldn't believe that he cared so much to do it for us'" (p. 356) as exemplification. According to the authors, students "perceived the comments as the evidence of the instructor's commitment to their progress" and that this and one other theme "emerged in the utterances of every participant of this group" (*ibid*).

Therefore, apparently, simply the act of providing individual comments can have a tonic effect, one which would likely be amplified when students are accustomed to receiving low amounts of such feedback. However, a small number of interviewees also explained how particularly their motivation and self-confidence were positively affected by marks or grades in cases where they had performed better than their peers. IU52, for example, stated:

> If you give me a higher mark than the other students, I will feel that I did better than others.

Indeed, the overall positive effect participants reported the feedback they received to have on their confidence and motivation was, in light of the supremacy of marks and grades in terms of perceived frequency, an unexpected finding. Koh (2008) maintained that "Assessment that promotes motivation needs to be perceived as a learning opportunity, not how well students perform represented by grades or marks achieved" (p. 225).

Reflecting this argument is a number of important reviews which concluded that students' motivation is more susceptible to being undermined in high-stakes-assessment environments where marks or grades are the primary types of feedback. Black & Wiliam (1998a), for example, stated that "feedback which draws attention away from the task and towards self-esteem can have a negative effect on attitudes and performance" (p. 23). By the latter form of feedback, the authors were referring to ego-involving feedback which, according to Butler (1987) and as discussed in Section 2.6, takes the form of grades and praise. James (2000) found this to be true in the case of mature students. Harlen & Deakin Crick (2003) also repeatedly called attention to the detrimental effect that grades can have on students' self-esteem, though this may vary depending on a

number of other factors, such as age, sex and whether students are high or low achievers.

One aspect which could help to explain the discrepancy between the study's finding and the literature is the heavily cohort-referenced system of assessment which appears to exist at the host institution. In such a system of assessment, knowledge of where students stand in relation to their peers is very important. This is reflected by IU52's comment above but also by QU168's statement:

> Feedback is basically an essential access to get a criterion about how you're doing your work because except for those extraordinary ones, *most students need to know whether they keep pace with others or not* (emphasis added).

In addition, when the interviewees were asked whether the feedback they received typically compared them to a set of criteria, their peers or themselves, 26 (86.7%) cited the second method and always through a ranking system. In contrast, nearly all interviewees stated that they preferred the first or third method. IU47, for example, explained that a ranking system had little meaning, particularly when the students in one group entered the course with different levels of expertise in that subject. He likened the process of learning to climbing a mountain and students with higher levels of expertise as starting the ascension from a higher altitude. Some would inevitably always be in front of others and those who compared themselves to their peers rather than to themselves and their own improvement would feel frustrated.

This is consistent with Buhagiar (2007) who pointed to the negative emotional impact of norm-referenced assessment on students. He argued that a criterion-referenced system of assessment – whereby students are judged according to a set of global criteria – or a

self-referenced system of assessment – whereby students' improvement is measured according to whether they have improved relative to previous performances – are likely to be most supportive of both learning and positives outcomes in terms of motivation. According to the author, these two referencing systems allow learners to build a "personal profile" of themselves – one that is reflective of individual or "absolute" (rather than relative) progress – and they foster greater cooperation among students due to reduced competition between them.

The interview schedule also included an item exploring the utility of praising and critical feedback in students' views. Their answers were almost unanimous. They reported that praising feedback generally led to an increase in their motivation and self-confidence; critical feedback, while potentially having a dampening effect on their self-esteem, better helped them to find or understand their faults and improve their learning. IU26, for example, explained:

> Although critical feedback can upset as well, it's more a kind of an alert to me, and tells me or urges me to find out the problems.

IU19 used similar wording and described that, while teachers "may use strong words" that made her "feel bad," they acted as an "alarm" which "urges" her to look more closely at her work. Similarly, IU39 explained that critical feedback helped her avoid repeating the same mistakes and added:

> Although sometimes it can be harsh, the point is to remember all criticism is for my best and I only need to extract the information that is helpful for me.

This runs counter to both Crooks' (1988) and Koh's (2008) recommendations that negative comments are best avoided due, in particular, to their motivationally dampening

effect. The participants appeared to see the value in critical feedback for the improvement of their learning, even if it was at the expense of their emotional wellbeing. Yang & Carless (2013) also noted that "Critical feedback can often be the most penetrating and useful" (p. 290). Hattie & Timperley (2007) argued that the effect of both positive and negative feedback on a particular learner depends on which level it is given as well as on their level of self-efficacy. Jonsson (2012) admitted that, while students prefer positive comments, these "have been shown to lead to less change […], and in order to improve, students also need critical comments" (p. 68). Indeed, a large proportion of respondents expected feedback to help them improve (see Section 4.3.5), and so greater acceptance of negative feedback may, according to this logic, not be surprising.

4.3.4 Timing of Feedback

The fourth and fifth items in the questionnaire investigated the timing of feedback: when the participants actually typically received it and when they preferred to receive it. Figure 4.8 provides a comparative view of the total number of responses for each timing category.

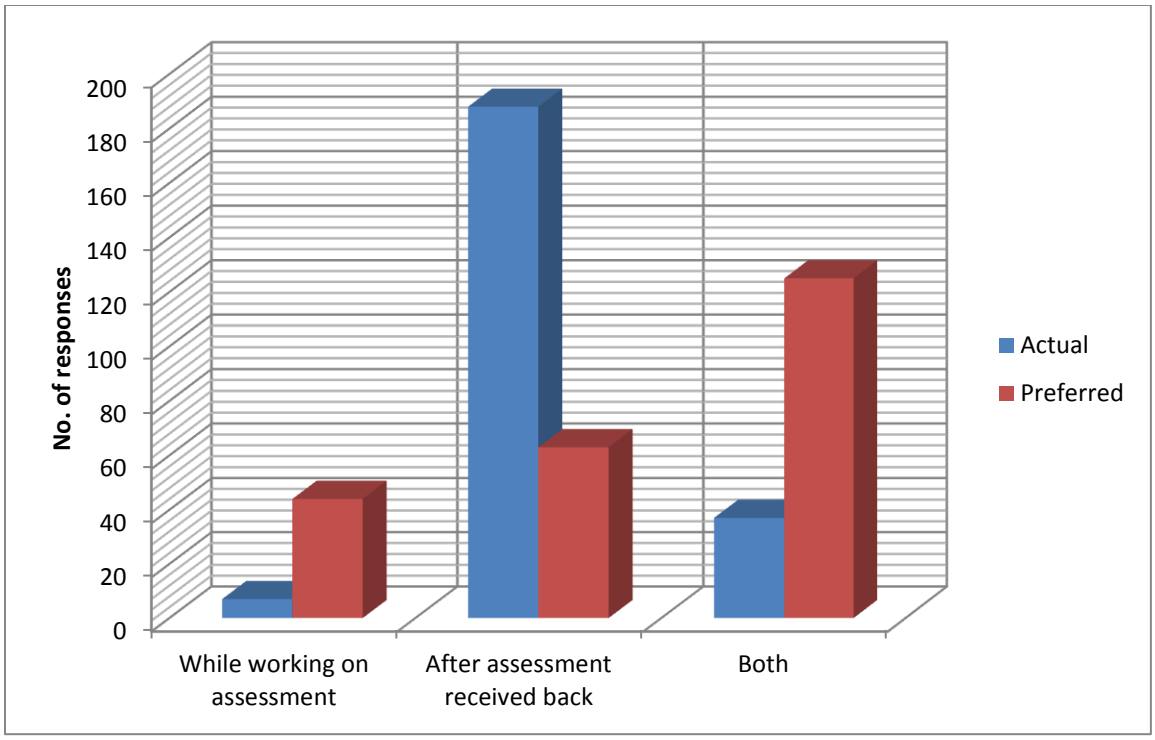


Figure 4.8 Actual vs. preferred timing of feedback

The results show a clear trend. The participants who believed they obtained feedback, generally, only after an assessment had been returned to them represents a substantial majority (81%). A very small number (3%) answered that they believed they usually received feedback only while they were working on an assessment with a higher though still relatively small number (16%) stating that they felt they received feedback both while working on an assessment as well as after turning it in.

Students' preferences contrast distinctly with their perceptions of the real situation. The majority (53.9%) answered that they favoured receiving feedback both during the process of working on an assessment and after turning it in. However, not inconsequential numbers of respondents stated that they preferred to receive feedback either only after an assessment had been returned to them (27.1%) or only while they were working on an assessment (19%).

Figure 4.8 clearly demonstrates the discrepancy between the respondents' experiences and preferences concerning the timing of feedback. While the incongruity is quite stark in all three categories, it is, at least in numerical terms, most extreme in the case of receiving feedback only after an assessment has been returned; far fewer students claimed they preferred this timing than was believed to actually occur.

Timing was, with references made in 35 (15.4%) answers, the most frequently occurring aspect amongst participants' responses to the seventh item (which asked respondents to describe the situation when feedback was most helpful for them) in the questionnaire. This reflects Mason & Bruning's (2001) positioning of this aspect as the key component (amongst six others) of learners' decision-making framework for the provision of feedback. It also mirrors the importance given to timing in terms of researchers' characterisation of "effective feedback" (Poulos & Mahony, 2008). Some examples of students' comments within this theme include:

> I receive it during working on a complicated assignment rather than after having finished it. For example, to make a computer programme, if the mistakes are found out too late, any kinds of debugging will be meaningless. In a word, timeliness of feedback is the most important factor which matters (QIU134).
>
> it's given to me when I'm trying to solve a complicated problem and have spent much time without getting any progress. I also consider it valuable after I've finished a project / an assessment, etc. (QU140).
>
> I finish the assessment, that is, when I have already thought by myself, then I have received the feedback. It helps me to compare (QU195).

It is interesting to note that QIU134 and QU140 preferred to receive feedback while completing a difficult task. This contrasts with Mason & Bruning's (2001) argument that students completing tasks with a high level of complexity ought to receive delayed feedback since this may give them the time necessary to "draw from [and 'actively process'] their previous information base and rethink incorrect information" (para. 31).

According to their responses, these two participants' preference for synchronous feedback is more likely explained by a perceived higher risk of completing the task incorrectly, leading to loss of time. In a similar vein, QU169 pointed out:

> If teachers could give me comments while I'm working on the assessment, time wouldn't be spent too much meaninglessly.

Timing was also investigated in the interviews. QIU30 pointed out that, since most of the assessments she did were quizzes and exams, she usually received feedback only after completion of the assessment. However, in the case of other types of assessments, such as essays and projects where multiple drafts were possible, she preferred to receive feedback during the assessment process. Similar to other students, IU16 stated that receiving critical feedback on such assignments during the assessment process was vital since, otherwise, he "may do a lot of work in vain" which would be "a real waste of time." QIU6 differentiated between different types and timing of feedback. For instance, while she received marks or grades after finalising an assessment, her teachers may provide the class or groups with oral comments beforehand.

Most students questioned on the issue of typically receiving feedback only after submitting an assessment felt the reason to be the large class sizes and the corresponding burdensome workload of the teachers. QIU119 reported that even the teaching assistants were too heavily engaged in other duties to be able to provide more than a score. IU156 concurred and added that this feedback was not provided in person but typically through the Internet, as this was more efficient. IU39 stated that more individualised feedback could be provided on assessments – for instance, on a draft of an essay – by visiting the teacher during their office hours. She admitted, however, that she and most of her peers rarely bothered doing so, again because of the limited time

made available by the teacher for such meetings and the high demand for the same. An additional explanation may be fear of the teachers, akin to IP167's remark cited in Section 4.3.3.

As noted above, more than a quarter of respondents answered that they actually favoured receiving feedback only after completing an assessment. Some of these were interviewed to uncover the possible reasons for this. IU26 explained that feedback given during the assessment process could be “disturbing” and cause him to “lose focus,” particularly when he was close to finding an answer by himself. This reminds of Scriven’s (1967) early caution of the intrusion of evaluation into the learning process. IU104 and IU94 appeared to stress independent learning and both stated that they would accept feedback while they were completing a task, but only if they actively sought it.

QIU18 explained that he preferred to receive feedback on a task once it had been completed since incomplete work was not “evidence” or a “valid” representation of his true ability. However, in this case, it is likely he was referring to summative judgments rather than formative comments, and not wanting the former type of feedback to be given prematurely. QIU113 portrayed any given task as learning and problem-solving processes which he preferred to complete unaided. He stated that, if he was interrupted during this process, he would “just follow the ideas of the professors,” leading to the production of a piece of work which was not wholly his own.

4.3.5 Feedback as an Enabler of Improvement

The numerically most prevalent theme emerging from any of the open-response items in the questionnaire was improvement. This was particularly the case with the sixth item

(which asked respondents to state what they considered to be the most important purpose of feedback). Of the 230 responses received to this item, 102 (44.4%) included the word "improve" or inflections thereof. A search using NVivo for the word "improve" including stemmed words and synonyms produced a total count of 161 references – by far, the highest value amongst any of the words. Some example responses manifesting this theme include:

> to make us know where and how can we improve ourselves, and then do better (QU50).
>
> to help us to identify our abilities and show us where we need improving (QU115).
>
> to tell you where you hadn't done well and what you need to improve and the most important is let you know how to improve (QIU119).

In an effort to triangulate this highly recurrent theme, a text frequency search using the "Find and Replace" tool in Microsoft Word instead was performed manually for synonyms (according to both *Cambridge Dictionaries Online* and *Merriam-Webster Online Dictionary*) and inflections of "improve." The search produced the following results: "help" (65 references), "better" (49 references), "correct" (25 references), "progress" (7 references), and "enhance" (2 references), totalling 148 references. If the words "improve" and "improvement" were included, the total number of references would even exceed those calculated by NVivo.

These figures help to confirm the finding that a large number of participating students expected the feedback they received to have a developmental or formative and also positive impact on their learning or learning experience. In the following are some examples of responses including some of the synonyms listed above:

> to tell students what they need to improve and motivate them to do better the next time (QU1).
>
> to help a student have deeper understanding of a concept and have more interest in thinking and solving a problem (QIU15).
>
> to give you some advice and let you make some progress (QU22).

Many students' responses mirror statements quoted in this thesis made by prominent authors and proponents of formative assessment and Assessment for Learning. QU217 simply wrote: "to improve." Some other examples, where respondents additionally contrasted formative assessment with summative assessment, include:

> to help me to figure out in which way I can do better, rather than just letting me know whether I have done well or poor (QU125).
>
> to tell students how to do better, rather than a ranking (QU148).
>
> to help students improve their ability not just use the grades to judge one person's ability (QU184).

Improvement also emerged as a major theme amongst students' responses to the seventh item (which asked respondents to describe the situation when feedback was most helpful for them) in the questionnaire. Of the 228 responses received to this item, 24 (10.5%) included the word "improve" or inflections thereof. An expanded search through NVivo including stemmed words and synonyms produced a total of 39 references. Some examples include:

> it is mainly about how to make improvements. I welcome criticism but it is always easier to criticize than to construct. So I'd prefer teachers' advice on how to revise the shortcomings and make the original work into a better one (QIU6).
>
> my teachers take it very seriously and try their best to help me analyse how to improve (QU90).
>
> I am confused with the reason why I did not do well, and I do not know how to improve (QU211).

These comments relate to a variety of aspects that the participants expected feedback to improve, primarily their academic performance and cognitive abilities. One other specific aspect emerged within the broader theme of improvement: studying habits. References to this aspect were made by 29 (12.6%) different students in response to the sixth item (which asked respondents to state what they considered to be the most important purpose of feedback) in the questionnaire. Two related aspects – learning skills and problem-solving abilities – were included in this category. Some examples include:

> to help us improve our learning abilities, such as acquiring different kinds of knowledge and solving problems by ourselves (QU17).
>
> to reveal my deficiency in logics and reasoning and providing valuable resources for remedy (QIU113).
>
> for the teachers to pinpoint the weak points or problems in studying which are shown in the assessment and to make students know directly so that they can recognize and overcome them (QU173).

With specific references made in 25 (11%) responses, studying habits also featured prominently in students' responses to the seventh item (which asked respondents to describe when feedback was most helpful for them) in the questionnaire. Some examples include:

> I could not find appropriate methods on how to study and what I should focus on (QU28).
>
> I learn something well and don't know how to do better (QU110).
>
> it tells me what I should change in the future in order to learn the lessons better (QU174).

QU86's response to the sixth item also highlights studying skills but makes an additional point that runs counter to the centrality of the acquisition of knowledge typical of Chinese

educational culture discussed in Section 1.3. According to her, the most important purpose of feedback was:

> to help students to improve their ability to study in the best way. How to study is more important than the knowledge learned in the university.

This reflects Ferguson's (2011) statement in Section 2.7 which underscores the important role that feedback can play in the development of certain skills whose application extends beyond academia into the professional sphere. The third key function Shute (2008) attributed to formative feedback, as discussed in Section 2.8, is also related to the improvement of studying skills, that is, learning methods.

It is recognised that the explicit and frequent mention of the words "improve" and "improvement" in the covering letter as well as questionnaire may have influenced participants in how they responded to this item and which language they used. Nonetheless, even allowing for a certain amount of bias, this finding is very important, placing it as a further centrepiece of what constitutes feedback for learning in the minds of the participants.

In addition, it mirrors the central feature that has been widely ascribed to both formative assessment and (formative) feedback: a closing or narrowing of the disparity between the actual level of performance and the desired learning goal (Ramaprasad, 1983; Sadler, 1989; Black & Wiliam 1998a; Draper, 2009; Walker, 2009; Wiliam, 2011). The centrality of this theme is also reflected in a synthesis from the Assessment Reform Group of lessons learned from the Analysis and Review of Innovations in Assessment (ARIA) project which encompassed a number of initiatives and developments in assessment across the U.K. In this pamphlet, ten "Principles of Assessment Practice"

are outlined, the first of which is that "Assessment of any kind should ultimately *improve* learning" (emphasis added, Gardner *et al*., 2008, p. 16).

4.3.6 Views over Reliance on Grades

Another theme which became apparent through the analysis of students' responses to the eighth item (which allowed respondents to offer additional comments) in the questionnaire was a perception of an overreliance on grades and, in some cases, also a related desire for more comments. A total of 27 responses (16.9%) can be attributed to this theme, making it, numerically, the most important amongst the three open-response items in the questionnaire after the much broader theme "improvement." Some examples of responses in this category include:

> In China, scores are widely used to evaluate students. Using scores can easily sort students in different grades, but it can hardly tell students where they can do better (QU44).
>
> Not just give comments of what is right or wrong. Give suggestions on what to do in the future. More written comments. Not just give a mark (QU46).
>
> I think in China people pay too much attention to scores, which makes students become utilitarian (QU151).

The latter comment is analogous to Bailey & Garner's (2010) statement that, "in a system in which turnover is high and the focus is on assessment, students become instrumentally motivated, focusing on marks rather than the educational value of written comments" (p. 189). Some students were particularly emphatic about their disenchantment with grades:

> It [(feedback)] is important, I think. BUT I DON'T WANT IT TO BE ALWAYS MARKS (emphasis in original, QU29).
>
> Grade is so cruel (QU48).

> In Chinese University, feedback always means mark, and only mark (QU60).

Other students took a more acute stance against grades:

> In my opinion, I think only a grade or just a mark that only shows the teacher had read your work make little sense (Chinese teacher sometimes write "阅" [*yuè*] which means "has looked" on your homework) (QIU63).
>
> Teachers always use marks and grades as feedback. It is easy for teachers but not helpful for students. Marks mean nothing. The only thing I care about is the suggestions from teachers (QU116).
>
> I don't think feedback should usually be given in the form of grades. Feedback is not a tool to divide students into various ranks (QU177).

Others stated a desire more generally for more detailed feedback. While QU122 expressed the wish "More often, more details," QU44 commented:

> I would like to have some comments as a feedback from the teacher. But I never had one in the last two years.

Ramsden (2003) argued that, irrespective of the mode of assessment, reporting results only in the form of marks and grades, is tantamount to "defrauding students" and is "unprofessional teaching behaviour [which] we ought not to tolerate" (p. 187). He described students as being "understandably angry" in these cases (*ibid*). However, in 13 (8.1%) responses to the eighth item, students showed understanding toward teachers' dependency on grades and inadequate quantity of comments.

Ramsden conjectured that instructors' unwillingness to provide feedback in other forms may be attributed to:

> Fear of losing one's authority by revealing the reasons for low marks; a mistaken notion that providing students with feedback is somehow helping the dull ones

> more than they deserve; [or] sheer laziness about making the effort to compose model answers or meet students (*ibid*).

Of course, other factors could (also) be to blame, for instance absent or insufficient initial teacher training or continuing professional development which many authors have argued is central in effectively applying formative assessment practices (e.g., Stiggins, 2002; Hattie, 2009; Bennett, 2011). Wiliam (2006b) wrote that "the task of improving formative assessment is substantially, if not mainly, about teacher professional development" (p. 287). The majority of the 13 students in question, however, cited more practical issues as being the main culprits: large classes and instructors' extraprofessional responsibilities, in particular their research and pressure to publish papers. Some example responses which reflect this include:

> Because there are usually many students in one class, it is a burden for teachers to make comments (QU49).

> As far as I can see, students in [HI] are absolutely lack of feedback. The reason may be there are too many students but not enough teachers - in many classes a teacher has to face 100-300 students. To make it worse, besides classes, usually a teacher has some other projects to work on, which means that he (or she) cannot focus on teaching totally (QU76).

> I feel so sorry and ashamed to say it, but here in [HI], I have never received any comments on my assignments from any teacher except English teachers and teaching assistant. You know, in classes relating to our own major, there are always hundreds of students in one class, and each teacher may be responsible for several classes. Therefore, it's impossible for teachers to correct and write comments on every student's assignment (QU79).

The picture these students painted reminds of Boud & Molloy's (2013) remark that "The practical dilemma of higher education is that the amount and type of feedback that can realistically be given is severely limited by resource constraints" (p. 703). This is likely all the more the case in China where, as described in Section 1.3, educational resources are stretched out between much larger student populations. In an examination into 31 secondary students' experiences of Assessment for Learning and feedback preferences

at a school in New Zealand, Cowie (2005) – through individual and group interviews – also found that "Some of the pupils appreciated that it was not possible for the teacher to provide feedback to each pupil on a one-to-one basis" (p. 150). The attitudes of these students and of those quoted above contrast with those included in a questionnaire survey conducted by Carless (2006) in which 1,740 students as well as 460 staff from eight universities in Hong Kong were asked about varying facets of assessment purposes and experiences. The author felt that the students "were not particularly sensitive to the workload implications for staff" (p. 226).

In her response to the eighth item, QU21 appeared to explain how she perceived obtaining feedback as her responsibility:

> We as students should forge the habit to contact more frequently with teachers to get feedbacks from them.

QU41 went one step further by arguing that students should not rely on feedback and adding:

> Actually it's more important to learn to judge (accurately) by myself. A student good enough should have this ability.

IU26 appeared to share this sentiment and pointed out that, once students left the academic environment and entered the professional sphere, they "should have the ability to judge themselves" and to know how to progress independently of external help since there would be fewer people to offer guidance and feedback. The attitude seemingly shared by these three students is strongly reminiscent of Hattie & Timperley's (2007) feedback about self-regulation (see Section 2.6) since, on this level, learners

engage in self-assessment and decide themselves if they want to solicit external feedback. Moreover, given the educational assessment climate that appears to prevail in China (see Sections 1.3 & 2.9.1), these views may be regarded as quite radical.

Other answers also remind of many of the elements which characterise this feedback type, including autonomy, self-control, self-direction and self-discipline. For example, QU158's response – which, at 268 words, was the longest received to any item – underscores teachers' willingness to provide feedback if it is sought. Therein, he pointed out:

> If you are willing to ask questions to teachers or assistants, they will surely answer you patiently.

Thereupon, he added: "But if you don't, no one will care," likely implying that students needed to take the first step in obtaining feedback. Another response corroborates this idea:

> Basically I think Chinese teachers won't give you many feedbacks until you are very active yourself. If you take an initiative to ask for advice, they'll be quite willing to help you. But if you are very shy and keep your mouth closed, then they don't know that you need help and all in all you'll only get a score at the end of the semester (QU70).

However, not all students seemed to be aware of this. QU69, for example, commented in her reply to the eighth item:

> After many years of study, we realized that students should ask for feedback.

QIU113 also explained that it was common practice to complete assignments independently and in quick succession, commenting that:

> In China, most students are not likely to communicate with their professors; they're accustomed to handing in their work and getting a grade, and being happy or sad, and doing it again.

Thus, it appears that, while a low incidence of comments can be partially accounted for by practical issues, such as large class sizes and instructors' other commitments, another reason may be some students' unawareness or unwillingness to take the initiative of seeking it on their own accord.

4.3.7 Feedback as a Means of Communication

Covering 10 (6.3%) responses, one further theme which emerged through the analysis of participants' answers to the eighth item (which allowed respondents to offer additional comments) in the questionnaire concerns the opportunity that feedback presents as an avenue of interaction or communication between students and teachers. Mory (2004) maintained that, in fact, the principal function of feedback as an enabler of knowledge acquisition is communication. Carless (2006) underscored the communicative importance of dialogic feedback or "assessment dialogues" in terms of their potential for helping to allay students' doubts and misunderstandings concerning their tutors as well as assessment processes.

Similarly, Nicol & Macfarlane-Dick (2006), as part of encouraging dialogue between students and their teachers as well as peers – another one of their seven principles of good feedback practice – argued that "Feedback as dialogue means that the student not only receives initial feedback information, but also has the opportunity to engage the teacher in discussion about that feedback" (p. 210). They admitted that this would be more difficult in large classes, particularly with respect to student-teacher dialogue. In this case, the authors suggested dividing students into small groups to discuss the

feedback they had received on their written assignments, or to collate their responses to questions posed in class and present them visually in the form of a histogram. However, the former idea would obviously not work well – regardless of the setting – if students, such as those who participated in this study, receive little written feedback; the latter idea would require significant additional effort to transcribe the responses and would necessitate the availability of certain hardware, software and computer knowhow.

QU157 noted that not only did feedback provide a good opportunity for facilitating improvement, but it was also "actually a collision of ideas, a kind of important communication." Further examples of responses in this category include:

> Maybe in-class feedback should increases. The situation now is that we nearly have no communication with the teacher during the class and in fact only a few students will use e-mail or phone calls to ask teacher questions (QIU15).
>
> I think feedback should be a kind of interaction. Teachers should ask students to feedback about teachers' feedback. Because students may not quite understand exactly teachers' comments on their work (QU35).
>
> In my opinion, feedback is something with whose help we can communicate with our professors. If both professors and us can use this method correctly, I'm sure we can get improvements in a short time (QU141).

Despite only covering nine (3.9%) responses, this theme also appeared in students' responses to the sixth item (which asked respondents to state what they considered to be the most important purpose of feedback) in the questionnaire, some of which are included here for purposes of added illustration:

> to show the student that the teacher actually cares him/her as an individual and is willing to communicate/interact with him/her (QU154).
>
> in order to get more improvement for my study and the abilities for communication skills with teachers (QP166).

> to give more instruction about how to improve my learning or to pick up more knowledge. I think it will also be okay for the feedback to be a connection between students and teacher, because sometimes students just like to get along with teachers like friends, which is a more casually communication atmosphere (QU230).

QP166, who in her answer to the eighth item (which allowed respondents to offer additional comments) in the questionnaire commented that feedback was important not only for students but also for teachers, further underscored the communicative value of feedback. However, upon being asked to expand on this point in the interview, she went one step further by arguing that feedback should occur iteratively; that is to say, once a student had been given feedback, they should feed back to the instructor with further questions or comments, which can "make the teacher know what the students are thinking." It is these feedback cycles which she considered as "communication" between students and teachers. This is strongly reminiscent not only of Hattie & Timperley's (2007) concept of "feed forward" discussed in Section 2.6, but also of Beaumont, O'Doherty & Shannon's (2011) dialogic feedback cycle which stresses the iterative nature of the dialogue between students and tutors as well as between peers.

This theme is important in that it is not a preferred characteristic attributed by students to feedback encountered in the literature. In view of the status of instructors relative to that of students in the Chinese education system as described in Section 1.3, and considering the related fear or lack of confidence many Chinese students seem to have of approaching their teachers, as alluded to in Sections 4.3.3 and 4.3.4, it may be surprising that any respondents expressed a desire for closer interaction with their teachers in matters of assessment. However, it is likely precisely because of these barriers that there appears to be at least a latent demand from the participants for more

direct communication with their teachers and for greater participation in the assessment process.

Drawing on Carless' (2011a) comments about Chinese students' passiveness in class, Black (2015) deduced: "It follows that formative assessment through oral dialogue *cannot* be established" (emphasis added, p. 168). In light of students' comments above, this statement may, however, be erroneous. Evidence gathered through this study suggests that formative assessment *can* be established through oral dialogue in China if: first, teachers engender a feeling of approachability in their students; second, students are aware of opportunities to approach their teachers or that they may be expected to take the initiative of seeking feedback from them; and, third, teachers' time resources are sufficient to meet the demand.

4.4 Greater Recognition as Individuals

In addition to the above themes, another underlying issue – closely associated with one of the aspects identified in Section 4.3.3 and that in Section 4.3.7 – may be said to have emerged: a desire on the part of the students to be recognised as individuals. Although not frequently occurring in the answers to a particular question in the questionnaire or interview schedule, it did surface in at least two of the open-ended items in the questionnaire as well as in some of the interviews.

Most of these instances, that is, related comments from students, have already been quoted throughout this chapter. For example, when explaining what they considered to be the most important purpose of feedback, QU25, QU31, QU116 and QU154 all expressed a wish for their teachers to show awareness of them individually. QIU8, IU108 and IU81 also stressed the emotional value of receiving attention from their teachers

through individualised feedback. In her additional comments, QU102 remarked on receiving less individual attention from her university than her high school teachers; IU156 did the same. A reference to this overarching theme can also be discerned in QU75's clarification of when feedback was most helpful for her: “I have a chance to communicate with teachers, I can tell opinions *on my own*” (emphasis added).

On the one hand, some students’ yearning for greater recognition as individuals may be surprising considering the collectivist character of Chinese society (see Section 2.6). On the other hand, this phenomenon may be expected if it is to be considered as a fundamental human trait that transcends cultural and historical factors. This question will not be pursued here since it reaches far beyond the thematic bounds of the study. Nonetheless, instructors’ ability to communicate more frequently with students individually and the provision of more individualised, in particular written, feedback being stymied especially by large class sizes is a serious problem.

One potential approach to alleviate this issue may be the use of electronic information technology for student-teacher communication and the provision of tutor feedback, for instance in written form through a web-based system or e-mail, or in audiovisual form through a chat video or screen-casting service. According to Lipnevich & Smith (2009), these methods have already become “an ingrained part of modern instructional practices,” one of the primary functions of which is to provide students feedback on their performance (p. 351). In a frequently cited book on this subject, Irons (2008) endorsed the use of information communication technologies to facilitate formative assessment and formative feedback, particularly in higher education and both in the case of classroom-based learning and distance learning. The author acknowledged staff training and a reliable technical infrastructure as prerequisites to the effective use of information

communication technologies, but averred that they "can reduce the workload for academic staff and potentially remove some of the constraints that make formative assessment and formative feedback impractical" (p. 91).

Some of the pedagogic benefits of these technologies Irons listed include greater communicative efficiency and flexibility for both students and teachers, less time spent marking, expedited provision of feedback, and greater scope for teachers to monitor students' progress. One further advantage, which encompasses these benefits, is increased individualised communication between students and their teachers as well as more frequent provision of written feedback – precisely those elements which the present study uncovered to be deficient at the host institution. A web-based system exists at the host institution, though it is not known to what extent it is used for these purposes or for purposes other than for the communication of students' grades and rankings. It is known, however, that e-mail is a widely available and commonly used facility at the host institution by both instructors and students, each of whom has an university e-mail account.

There is ample of evidence to suggest that e-mail can be a valuable pedagogical tool in a number of different settings. For example, Yu & Yu (2002), in an investigation into the impact of e-mail on the achievement and perceptions of a group of Taiwanese trainee teachers, found empirical evidence suggesting that it helped to promote their cognitive growth. In a review of literature on the use of e-mail as a means to provide feedback, Huett (2004) deduced that "The ubiquitous use of email for feedback in the classroom is lending the medium a new level of credence as an educational tool" (para. 42). In a study exploring the effects of using e-mail to improve English reading skills on three

groups of Iranian high school students, Taki & Ramazani (2011), found the impact to be "statistically significant" (para. 1).

Despite this evidence, Budge (2011), drawing on the results of her examination of an Australian university cohort's perceptions of electronic feedback, pointed out that "there is a human aspect to feedback that is conveyed through non-electronic forms that students value very highly" (p. 348). Thus, it may be prudent that e-mail and perhaps other electronic media be given supportive roles rather than being relied on to fill the communication and feedback gaps.

4.5 Conclusions

All of the research questions the study set out to investigate were answered. With the exception of comments written at the end of their assignments, the participating students appeared to hold a preference for written comments. Individual oral comments were slightly less popular but those given collectively to groups or the whole class during lessons were generally deemed ineffective. While not amongst the most popular forms of feedback, many students still regarded marks or grades as useful in supporting their learning endeavours.

This type of feedback was also perceived as being the most frequently given. Oral comments given collectively to groups or the whole class during lessons was the only other form of feedback which, on average, was rated as frequently given. All other types were perceived to be provided only infrequently. While several respondents held an aversion to the high frequency with which they believed to receive feedback in the form of marks or grades, others appeared to be sympathetic and valued the little individual attention from their teachers they did receive. Several also pointed out that individual

feedback was available upon request, given that they were able to make an appointment with the relevant instructor.

The findings also indicate a largely positive effect of the feedback students received on all four aspects included in the questionnaire: their ability to achieve better results on their assessments, their ability to improve their studying habits, their motivation and their self-confidence, in particular on the first aspect. The data also show that the vast majority of the respondents felt that they received feedback normally after they had completed an assignment while most preferred to be given feedback both while they were working on an assignment as well as after turning it in.

The participants expected feedback to fulfil a number of different functions. First and foremost, they expected it to support them in improving their learning, in particular their studying habits. The students also expressed a desire for feedback to play a supportive role in terms of affect, in particular: a sense of being paid attention to or approval from teachers; the holding or heightening of interest in the class content; and the safeguarding or reinforcement of self-confidence and motivation. In spite of other possibly contradictory findings, some students also called attention to the important function of feedback in enabling them to communicate with their instructors. Some participants also attached importance to feedback as an enabler of communication between them and their instructors and, thus, to give them greater recognition as individuals.

Many of the themes that appeared in students' responses to the questionnaire item investigating which functions they expected feedback to fulfil also appeared in that on the subject of what they believed to be important hallmarks of effective feedback. These

comprise the improvement of their learning, timing and a range of psychological aspects, including emotional support when students feel depressed or frustrated as well as self-confidence and motivation.

It is appropriate to conclude by briefly viewing the study's findings more holistically and considering how they fit into the many conceptions and models of formative assessment and feedback discussed in Chapter 2. The findings do not neatly fit into any particular framework, but this is unsurprising since the study did not intend to prove or disprove a specific theory. Rather, the ambition in was, where possible, to draw parallels between the results and other works, acting to validate or contest already existing findings and ideas – both those derived in Western contexts and those reported on in the limited evidence available in English in the Chinese context.

Signifying one of the greatest parallels between the literature (Crooks, 1988; Black & Wiliam 1998a; Carless 2006; Shute, 2008) and the study's findings is students' preference for written feedback and their criticism of feedback deficient in detail. Despite a relative lack of works on this subject reported by some authors, another congruence between the literature (Assessment Reform Group, 1999 & 2002; Shute, 2008; Värlander, 2008; Evans, 2013) and the study's results is participants' desire for feedback to play an emotionally supportive role. The most significant divergence between the study's findings and the literature (Butler, 1987; Crooks; 1988; Harlen & Deakin Crick, 2002 & 2003) is students' positive testimony regarding the effect of the feedback they received on their achievement and aspects of their psychology despite the hegemonic role of marks and grades in terms of perceived frequency.

Perhaps most notable in more general terms is the convergence between the participants' and scholars' (such as those quoted throughout Section 2.6) stance towards the central role of feedback in education and, more specifically, in improving learning. The large number of responses to this effect and remarks like "Irrespective of volume, feedback is an indispensable part of our learning" (QIU8) and "Feedback is essential in college!!!!" (QIU18) are testament to this. QU135's comment that "a good feedback should include: What are you doing now? What could you do in the next step? And it should be suggestions instead of orders" closely resembles the three questions Hattie & Timperley (2007) required for effective feedback to be able to answer (see Section 2.6) as well as the Assessment Reform Group's (2002) conceptualisation of Assessment for Learning (see Section 2.4).

In conclusion, it can be deduced that the study's participants were not only very opinionated on the topic of feedback but also keenly aware of its power. An important question raised by this deduction is whether it also applies to the majority of other Chinese students in China and elsewhere. If yes, it is imperative for educationists and policy-makers to listen to the opinions of the recipients of feedback, and to seize the promise that feedback holds for the improvement of learning. This and other appeals are made to researchers in the following, final chapter after a recapitulation of the major objectives of this thesis and a discussion of the study's aims, implications and limitations.

Chapter 5

Looking Back and Looking Forward

5.1 Synopsis

This thesis attempted to fulfil a number of objectives. Its first major goal was to depict the emergence, evolution and formalisation of formative assessment as one of the major themes within educational assessment. This was necessary to set the scene for the focus of the thesis: feedback. A closer examination of the concept "formative" was important due to its centrality within discussions on formative assessment and formative feedback, and to later aid in differentiating the latter from feedback for learning. A conceptualisation of feedback and a review of some of the most significant literary contributions on this topic were followed by focused discussions of this pedagogical tool within higher education. A second chief goal of this thesis was to paint a picture of formative assessment and feedback in mainland China, first by illuminating the stance, policies and initiatives of the Chinese Central Government and, second, by reviewing the relevant literature available in English specifically within this geographical context.

A further objective of this thesis was to elucidate all the key methodological elements necessary for providing a sound theoretical as well as practical foundation upon which to situate the study. It was characterised as a mixed-methods investigation and a justification was provided as to its emphasis on qualitative data. Issues related to bias, epistemology and the judging of the scientific rigor of a qualitatively focused study were reflected on. Consideration was also given to methodological issues native to China. Following a chronological depiction of the major steps involved in the preparation of the study were accounts of the pilot study, sampling and data analysis techniques, ethical

concerns, and an elaboration on and justification of the data collection methods used in the study.

The ambition of the study was to answer a number of questions, including: how effective participants found each type of feedback included in the questionnaire and how often they believed they received each one; how they felt the feedback they received generally affected their academic performance, studying habits, motivation and self-confidence; when they believed as well as preferred to receive feedback; which roles they expected feedback to play; and how they understood effective feedback. Participants generally expressed a preference for comments over marks or grades; for written comments over oral comments; and for individual oral comments over collective oral comments. Marks or grades and collective oral comments were perceived as being the most frequently provided feedback types while others were believed to be less – in some cases, far less – frequently given. Participants reported an overall positive effect of the feedback they received on all four aspects included in the study. A strong majority felt that they received feedback usually after they had completed an assignment whereas most preferred to be given feedback both while they were working on an assignment as well as after handing it in. Nearly half of participants wished feedback to aid in the improvement of their learning. The largest minorities expected feedback to play psychologically supportive roles and characterised effective feedback as fulfilling both these criteria.

5.2 Implications for Theory and Practice

There are several implications which emanate from the study's findings. An important consequence for teachers, and to some extend also for scholars, may be that students do not need convincing of the value of feedback. Instead, the focus ought to be on

bringing to fruition the benefits that learners are already aware feedback can have for them. If this outcome is not being produced, then the answer may lie with the manner in which the feedback is being communicated rather than (only) with the students themselves. For example, the participants expressed a clear preference for written feedback and individual oral feedback, and a very large number stated an expectation for it to aid in the improvement of learning outcomes and of various affective factors.

Instructors may argue that meeting these expectations is not feasible since providing more written comments and individual oral feedback would inevitably place even greater pressure on already limited resources. Without an increase in the latter variable, a possible solution would be to reduce the total amount of assessed work, and to provide more of the above two types of feedback and fewer marks or grades on the work which is assessed. While not discussed at length in this thesis and although perhaps more difficult to implement owing to cultural factors, peer- and self-assessment could also be given greater roles. These two initiatives may go some way to meeting some of the participants' other expectations, such as improving specificity, better supporting or protecting their self-confidence and motivation, increasing direct communication between them and their teachers, and giving them greater recognition as individuals.

In particular the second initiative, however, may require, in some cases, considerable modifications to instructors' teaching methodology and assessment practices. These could be realised through greater investment into initial teacher training and continuing professional development which could take the form of seminars and workshops. Considering the importance of assessment in education, attendance should be made mandatory. The completion of accredited and more formal, intensive courses could lead to a vocational qualification as an added incentive. Particularly in the case of continuing

professional development, however, teaching staff should be given temporary leave to avoid conflicts with their regular professional obligations.

Establishing formal lines of communication between the teaching faculty and student representatives would also be essential since it would, otherwise, be difficult to establish whether the teacher training and any policy changes were having the desired effect. Unannounced observations from senior staff members, meetings with instructors and interviews (if possible, meeting the same ethical requirements as those conducted during this study) with randomly selected students could also play an important role in ensuring that instructors were translating theory into practice. In addition, the evaluations (by students of their teachers), which are conducted electronically at the host institution twice annually, could also be given a greater role in helping to identify where assessment practices could be further improved.

Although these suggestions are, generally, sensible and would likely bear fruit in any context, a disclaimer is necessary. Any changes required of instructors in terms of their assessment practices, or otherwise, could be interpreted by them as an intrusion into their own professional domain and may require a considerable degree of open-mindedness on their part. This surely pertains to any context but perhaps all the more in ones where contrasting cultures coexist, such as at the host institution which employs a large number of both Asian and Western lecturers. To transplant assessment techniques, as they are typically practised in one setting into another, without making considerations for cultural, historical and socio-political factors would be imprudent. The same may apply to students who may suffer from “assessment culture shock” if they are too swiftly confronted with radically different assessment practices, even if these are well-intended.

The changes suggested above may also have more far-reaching effects. They may call for educational institutions to make adjustments to their curricula and to the criteria for entry, graduation, the awarding of scholarships and the like. This, in turn, may impact the national education system as a whole. The Chinese Central Government has long recognised the importance of formative assessment and has taken steps to implement it. For its efforts to bear fruit, closer collaboration between the Government and educational institutions, that between educational institutions and departments or teachers, and that between teachers and students is needed, each feeding up, back and forward to one another.

5.3 Limitations and Recommendations for Future Research

Before concluding this thesis, it is important to acknowledge the study's limitations and to make recommendations for similar research in the future. One issue concerns confirmability in the analysis of the written responses. Since the majority of answers were quite short and contained little detailed elaboration, it may not always have been apparent what the respondents' actual views were. This problem was exacerbated by the fact that they were writing in a foreign language. This led to some inappropriate word usage and other lexical, syntactic and grammatical inaccuracies which themselves may have caused misinterpretations. For the same reason, a compromise, particularly in the way motivation was conceptualised, also needed to be made in order to make the concept more accessible to the participants. This is important when comparing the findings of this study related to this aspect to those of others.

This was one of the motivations behind combining the questionnaire with an interview since the latter method allowed not only for deeper probing of a participant's views but also for the clarification of their questionnaire responses. Mitigating the second issue is

the researcher's considerable experience of working with Chinese learners of English as a Foreign/Second Language. During this time, he has become aware of errors typically made by these learners and of the intended meaning of statements made containing such errors. However, greater credibility and dependability would likely have been rendered if the questionnaire had been written and the interviews conducted in Chinese, even though translation of the data from Chinese into English would also have been consequential in this regard.

A further point is potential differences between instructors with respect to the type, frequency and timing of feedback. Students' experiences of feedback may have been biased through taking a particular lecturer's course(s) or more courses from one or several lecturers than from others whose feedback practices differ significantly. Differences between respondents could also be explained by their year of study. While second-, third- and fourth-year students may have had ample experience of receiving feedback, particularly first-year students may not yet have. In addition, while no noteworthy differences were found between those natural and social science students who were interviewed in terms of their impression of how frequently they received marks or grades and comments, a relationship may have existed within the larger sample.

The above factors may have led to a skewing of the data. The inclusion of additional questions in the questionnaire regarding the respondents' area as well as year of study and the subsequent differentiation of the data according to these aspects may have alleviated potential bias as well as enabled the making of potential inferences. Through the answer to one of the open-response questions, it also came to light that, in the case of larger classes, the correction of assignments was undertaken by a number of teaching

assistants. It is important, therefore, to acknowledge that participants' experiences of feedback also reflected, to some extent, the feedback practices of teaching assistants.

It also needs to be underscored that participants were asked to base their answers specifically on the experiences of feedback they had gathered at the host institution. As a result, their responses may not have reflected their beliefs in broader terms. For example, while a respondent stated that they found marks or grades an ineffective form of feedback as per their experiences at the host institution, they may have believed that it was effective in theory. Moreover, they may have changed their views after taking part in the study. Thus, it may have been more suitable to investigate participants' views without this geographic restriction, particularly since the investigation was not a case study. Moreover, investigating students' satisfaction with the frequency they perceived to receive certain feedback types rather than simply how often they believed to receive each one would have been truer to a qualitatively focused study.

One realisation that was made towards the completion of the interviews, which is particularly consequential for future work on feedback, is the need to more carefully consider the magnitude of this topic as a study focus. Feedback manifested itself as a subject with more layers and more facets than initially expected. Fewer aspects should have been examined and those aspects could have been broken down into smaller elements in order to achieve greater depth of understanding. The timing of feedback is a good example. It is or should not have simply been a matter of receiving feedback before, during or both before or during an assessment. Rather, and as pointed out in Section 4.3.4, different types of feedback can be received at different stages during the assessment process. Furthermore, there are different forms of assessment where one or several timings of feedback may not apply or be more or less applicable than others.

5.4 Final Thoughts

Taras (2008) observed: “Evaluation or assessment is a ubiquitous activity pervading our every thought, opinion and action. Life is unthinkable without the making of judgements to inform our decisions, either consciously or unconsciously” (p. 394). In China, where examinations were first employed to measure performance and where competition, particularly in the education sector, is now so fierce that passing an exam may be perceived by some students as a matter of life and death, this may be all the more the case.

Thus, in addition to satisfying academic curiosity and improving educational practices, better understanding assessment and its products from students’ perspective may, in some societies, assume a more fundamental or existential purpose. Governments, scholars and educationists have recognised the merit of using assessment not only to quantify but also advance learning. Initiatives in both the U.K. and China have been taken to translate theory into practice, most commonly under the heading of “formative assessment” or “Assessment for Learning.” As one of the components in the circuitry of these practices most proximate *vis-à-vis* the learners is feedback, more specifically formative feedback.

Like formative assessment and Assessment for Learning, the argument that feedback should be “gap-closing” has enjoyed widespread acknowledgement. To demonstrate how it may best serve educational needs, feedback has been dissected, categorised and modelled. It has been placed within theoretical multidimensional prisms to refract out those components which are most supportive of learning outcomes as well as to highlight from which reference points this will most likely be the case.

While this is undoubtedly a valuable undertaking, it is vital that research into feedback in higher education as well as other sectors be more momentously informed by the experiences and perspectives of students since they are the major stakeholders and foremost consumers of feedback. Moreover, it is imperative that more such work be conducted in non-Western and non-Anglophone settings in order to further enrich work in this area with a greater diversity of epistemologies and allow for interregional and cross-cultural comparisons to be made. This may go some way towards better understanding and forecasting the impacts of feedback on learners and, ultimately, towards conceiving a framework which allows for contextually sensitive methodologies of generating feedback which are not only formative but also for-learning.

References

Agee, J. (2009). Developing qualitative research questions: a reflective process, *International Journal of Qualitative Studies in Education*. 22 (4), pp. 431-447.

Arksey, H. & Knight, P. (1999). *Interviewing for social scientists*. London: Sage Publications.

Askew, S. (ed.) (2002). *Feedback for Learning*. London: Routledge/Falmer.

Assessment Reform Group (1999). *Assessment for Learning: Beyond the Black Box*. Cambridge: University of Cambridge School of Education.

Assessment Reform Group (2002). *Assessment for Learning: 10 Principles*. Cambridge: University of Cambridge School of Education.

Atkin, J. M. (1963). Some Evaluation Problems in a Course Content Improvement Project. *Journal of Research in Science Teaching*. 1 (2), pp. 129-132.

Bailey, R. & Garner, M. (2010). Is the feedback in higher education assessment worth the paper it is written on? Teachers' reflections on their practices. *Teaching in Higher Education*. 15 (2), pp. 187-198.

Bandura, A. (1977). Self-efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review*. 84 (2), pp. 191-215.

Bandura, A. (1981). Self-referent thought: A developmental analysis of self-efficacy. In: Flavell, J. H. & Ross, R. (eds.). *Social Cognitive Development: Frontiers and Possible Futures*. Cambridge: Cambridge University Press.

Bandura, A. (1997). *Self-efficacy: The Exercise of Control*. New York: W. H. Freeman.

Bandura, A. & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology*, 41 (3), pp. 586-598.

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A. & Morgan, M. T. (1991). The Instructional Effect of Feedback in Test-Like Events. *Review of Educational Research*. 61 (2), pp. 213-238.

Barnes, M., Clarke, D. & Stephens, M. (2000). Assessment: the engine of systemic curricular reform? *Curriculum Studies*. 32 (5), pp. 623-650.

Beaumont, C., O'Doherty, M. & Shannon, L. (2011). Reconceptualising assessment feedback: a key to improving student learning? *Studies in Higher Education*. 36 (6), pp. 671-687.

Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*. 18 (1), pp. 5-25.

Berry, R. (2011). Educational assessment in mainland China, Hong Kong and Taiwan. In: Berry, R. & Adamson, B. (eds.). *Assessment Reform in Education: Policy and Practice*. Dordrecht, Netherlands: Springer.

Biggs, J. (1998). Assessment and Classroom Learning: a role for summative assessment? *Assessment in Education: Principles, Policy & Practice*. 5 (1), pp. 103-110.

Black, P. (2015) Formative assessment – an optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice*. 22 (1), p. 161-177.

Black, P., Harrison, C., Hodgen, J., Marshall, B. & Wiliam, D. (2005). Dissemination and evaluation: a response to Smith and Gorard. *Research Intelligence*. Issue 93, p. 7.

Black, P., Harrison, C., Lee, C., Marshall, B., Wiliam, D. (2003). Assessment for Learning: Putting it into practice. Maidenhead: Open University Press.

Black, P. & Wiliam, D. (1998a). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*. 5 (1), pp. 7-71.

Black, P. & Wiliam, D. (1998b). *Inside the Black Box: Raising standards through classroom assessment*. London: University of London, King's College.

Black, P. J., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*. 21 (1), pp. 5-31.

Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. In: Taylor, R. W. (ed.). *Educational evaluation: New roles, new means: The Sixty-Eighth Yearbook of the National Society for the Study of Education, Part II*. Chicago: 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.

Bloom, B. S. & Sosniak, L. A. (1985). *Developing Talent in Young People*. New York: Ballantine Books.

Bloxham, S. & Campbell, L. (2010). Generating dialogue in assessment feedback: exploring the use of interactive cover sheets. *Assessment & Evaluation in Higher Education*. 35 (3), pp. 291-300.

Boud, D. & Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in Higher Education*. 38 (6), pp. 698-712.

Boyatzis, R. E. (1998). *Transforming Qualitative Information: Thematic Analysis and Code Development*. Thousand Oaks, California: Sage Publications.

Boyle, W. & Charles, M. (2010). Leading Learning through Assessment for Learning? *School Leadership & Management*. 30 (3), pp. 285-300.

Brinkmann, S. (2013). *Qualitative Interviewing: Understanding Qualitative Research*. New York: Oxford University Press.

British Educational Research Association (2011). *Ethical Guidelines for Educational Research*. London: BERA.

Broadfoot, P. (1993). Exploring the Forgotten Continent: A Traveller's Tale. *Scottish Educational Review*. 26 (2), pp. 88-96.

Brookhart, S. M. (2007). Expanding views about formative classroom assessment: a review of the literature. In: McMillan, J. H. (ed.) *Formative classroom assessment: theory into practice*. New York: Teachers College Press.

Brooks, V. & Fancourt, N. (2012). Is self-assessment in religious education unique? *British Journal of Religious Education*. 34 (2), pp. 123-137.

Brown, J. (2007). Feedback: the student perspective. *Research in Post-Compulsory Education*. 12 (1), pp. 33-51.

Bryman, A. (2006). Integrating quantitative and qualitative research: how is it done? *Qualitative Research*. 6 (1) pp. 97-113.

Bryman, A. (2012). *Social Research Methods*. 4th Ed. Oxford: Oxford University Press.

Budge, K. (2011). A Desire for the Personal: Student Perceptions of Electronic Feedback. *International Journal of Teaching and Learning in Higher Education*. 23 (3), pp. 342-349.

Buhagiar, M. A. (2007). Classroom assessment within the alternative assessment paradigm: revisiting the territory. *The Curriculum Journal*. 18 (1), pp. 39-56.

Burke, D. (2009). Strategies for using feedback students bring to higher education. *Assessment & Evaluation in Higher Education*. 34 (1), pp. 41-50.

Burke, J. R. & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*. 33 (7) pp. 14-26.

Butler, D. L. & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*. 65 (3), pp. 245-281.

Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology*. 79 (4), pp. 474-482.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*. 31 (2), pp. 219-233.

Carless, D. (2011a). From Testing to Productive Student Learning: Implementing Formative Assessment in Confucian-Heritage Settings. New York: Routledge.

Carless, D. R. (2011b). Reconfiguring assessment to promote productive student learning. In: Powell-Davies, P. (ed.). *New Directions: Assessment and Evaluation - A collection of papers* (East Asia). British Council. Available from: http://www.britishcouncil.jp/sites/britishcouncil.jp/files/eng-new-directions-en.pdf. [Accessed: 17 November 2014]

Carless, D. (2013). Sustainable feedback and the development of student self-evaluative capacities. In: Merry, S., Price, M., Carless, D. & Taras, M. (eds.). *Reconceptualising Feedback in Higher Education: Developing Dialogue with Students*. Abingdon: Routledge.

Carless, D. & Lam, R. (2014). Developing Assessment for Productive Learning in Confucian-Influenced Settings: Potentials and Challenges. In: Wyatt-Smith, C., Klenowski, V. & Colbert, P. (eds.). *Designing Assessment for Quality Learning*. Dordrecht: Springer.

Carless, D., Salter, D., Yang, M. & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education*. 36 (4), pp. 395-407.

Carson, J. G, & Nelson, G. L. (1996). Chinese Students' Perceptions of ESL Peer Response Group Interaction. *Journal of Second Language Writing*. 5 (1), pp. 1-19.

Central Intelligence Agency (2014). *The World Factbook: United Kingdom* [Online]. Available from: https://www.cia.gov/library/publications/the-world-factbook/geos/uk.html [Accessed: 17 November 2014]

Charmaz, K. (2000). Grounded Theory: Objectivist and Constructivist Methods. In: Denzin, N. K. & Y. S. Lincoln (eds.). *Handbook of Qualitative Research*. 2nd Ed. Thousand Oaks, California: Sage Publications.

Charmaz, K. (2014). *Constructing Grounded Theory*. 2nd Ed. London: Sage Publications.

Chelala, C. (2014). Society should address suicide among youths. *China Daily* [Online]. July 29th. Available from: http://www.chinadaily.com.cn/opinion/2014-07/29/content_17945717.htm. [Accessed: 17 November 2014]

Chen, Q., Kettle, M., Klenowski, V. & May, L. (2012). Interpretations of formative assessment in the teaching of English at two Chinese universities: a sociocultural perspective. *Assessment & Evaluation in Higher Education*. 38 (7), pp. 831-846.

Chen, Q. & Klenowski, V. (2009). Assessment and curriculum reform in China: EFL education at the tertiary level. *2008 AARE International Education Conference*. Brisbane, November 30th – December 4th 2008. Brisbane: Queensland University of Technology. Available from: http://www.aare.edu.au/data/publications/2008/che08574.pdf. [Accessed: 17 November 2014]

Chen, Q., May, L., Klenowski, V. & Kettle, M. (2014). The enactment of formative assessment in English language classrooms in two Chinese universities: teacher and student responses. *Assessment in Education: Principles, Policy & Practice*. 21 (3), pp. 271-285.

Cheng, L. (2008). The key to success: English language testing in China. *Language Testing*. 25 (1), pp. 15-37.

Cheng, L., Rogers, T. & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: purposes, methods, and procedures. *Language Testing*. 21 (3), pp. 360-389.

Cheung, H. Y. & Hui, S. K. F (2003). Mainland Immigrant and Hong Kong Local Students' Psychological Sense of School Membership. *Asia Pacific Education Review*. 4 (1), pp. 67-74.

China Education Center (2014). *Project 211 and 985* [Online]. Available from: http://www.chinaeducenter.com/en/cedu/ceduproject211.php. [Accessed: 17 November 2014]

Cho, J. & Trent, A. (2006). Validity in qualitative research revisited. *Qualitative Research*. 6 (3), pp. 319-340.

Cohen, L., Manion, L. & Morrison, K. (2011). *Research Methods in Education*. 7th Ed. London: Routledge.

Cooper, K. & White, R. E. (2012). *Qualitative Research in the Post-Modern Era*: *Contexts of Qualitative Research*. Dordrecht: Springer.

Cowie, B. (2005). Pupil commentary on assessment for learning. *The Curriculum Journal*. 16 (2), pp. 137-151.

Creswell, J. W. (1998). *Qualitative Inquiry And Research Design: Choosing Among Five Traditions*. Thousand Oaks, California: Sage Publications.

Creswell, J. W. (2003). *Research Design: Qualitative, Quantitative, and Mixed Method Approaches*. 2nd Ed. Thousand Oaks, California: Sage Publications.

Creswell, J. W. & Miller, D. L. (2000). Determining Validity in Qualitative Inquiry. *Theory Into Practice*. 39 (3), pp. 124-130.

Crooks, T. J. (1988). The Impact of Classroom Evaluation Practices on Students. *Review of Educational Research*. 5 (4), pp. 438-481.

Cross, K. P. (1998). Classroom Research: Implementing the Scholarship of Teaching. In: Angelo, T. (ed.). *Classroom Assessment and Research: An Update on Uses, Approaches, and Research Findings*. San Francisco, California: Jossey-Bass.

Crossouard, B. (2009): A sociocultural reflection on formative assessment and collaborative challenges in the states of Jersey. *Research Papers in Education*. 24 (1), pp. 77-93.

Delamont, S. (2002). *Fieldwork in educational settings: methods, pitfalls and perspectives*. London: Taylor and Francis.

Dello-Iacovo, B. (2009). Curriculum reform and 'Quality Education' in China: An overview. *International Journal of Educational Development*. 29 (3), pp. 241-249.

Denzin, N. K. (1970). *The Research Act in Sociology: A Theoretical Introduction to Sociological Methods*. London: Butterworths.

Denzin, N. K. & Lincoln, Y. S. (2005). The Discipline and Practice of Qualitative Research. In: Denzin, N. K. & Lincoln, Y. S. (eds.). *The Sage Handbook of Qualitative Research*. 3rd Ed. Thousand Oaks, California: Sage Publications.

Department for Children, Schools and Families (2008). *The Assessment for Learning Strategy*. Nottingham: DCSF Publications. Available from: http://dera.ioe.ac.uk/8161/1/DCSF-00341-2008.pdf. [Accessed: 17 November 2014]

Dowden, T., Pittaway, S., Yost, H. & McCarthy, R. (2013). Students' perceptions of written feedback in teacher education: ideally feedback is a continuing two-way communication that encourages progress. *Assessment & Evaluation in Higher Education*. 38 (3), pp. 349-362.

Draper, S. W. (2009). What are learners actually regulating when given feedback? *British Journal of Educational Technology*. 40 (2), pp. 306-315.

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), pp. 1-11.

Eimer, D. (2011). Wave of suicide sweeps China's graduate class. *The Telegraph* [Online]. Available from: http://www.telegraph.co.uk/news/worldnews/asia/china/5907368/Wave-of-suicide-sweeps-Chinas-graduate-class.html. [Accessed: 17 November 2014]

Elwood, J. & Klenowski, V. (2002) Creating Communities of Shared Practice: the challenges of assessment use in learning and teaching. *Assessment & Evaluation in Higher Education*. 27 (3), pp. 243-256.

Evans, C. (2013). Making Sense of Assessment Feedback in Higher Education. *Review of Educational Research*. 83 (1), pp. 70-120.

Ferguson, P. (2011). Student Perceptions of Quality Feedback in Teacher Education. *Assessment & Evaluation in Higher Education*. 36 (1), pp. 51-62.

Fielding, N. G. & Fielding, J. L. (1986). *Linking Data: Qualitative and Quantitative Methods in Social Research*. Beverley Hills, California: Sage Publications.

Flick, U. (2002). *An Introduction to Qualitative Research*. London: Sage Publications.

Fluckiger, J., Tixier y Vigil, Y., Pasco, R. & Danielson, K. (2010). Formative Feedback: Involving Students as Partners in Assessment to Enhance Learning. *College Teaching*. 58 (4), pp. 136-140.

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*. 12 (2), pp. 219-245.

Ford, D. Y., Moore, J., Whiting, G. & Grantham, T. (2008). Conducting Cross-Cultural Research: Controversy, Cautions, Concerns, and Considerations. *Roeper Review*. 30 (2), pp. 82-92.

Fox, J., Murray, C. & Warm, A. (2003). Conducting research using web-based questionnaires: Practical, methodological, and ethical considerations. *International Journal of Social Research Methodology*. 6 (2), pp. 167-180.

Fuchs, L. S. & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*. 53 (3), pp. 199-208.

Gamlem, S. M. & Smith, K. Student perceptions of classroom feedback. *Assessment in Education: Principles, Policy & Practice*. 20 (2), pp. 150-169.

Gardner, J., Harlen, W., Hayward, L. & Stobart, G. (2008). *Changing Assessment Practice: Process, Principles and Standards* [Online]. Assessment Reform Group. Available from: http://www.aria.qub.ac.uk/JG Changing Assment Practice Final Final.pdf. [Accessed: 27 March 2015]

Gibbs, G. & Simpson, C. (2004-2005). Conditions Under Which Assessment Supports Students' Learning. *Learning and Teaching in Higher Education*. 1 (1), pp. 3-31.

Gillham, B. (2007). *Developing a Questionnaire*. 2nd Ed. London: Continuum.

Gipps, C. (1994). Beyond Testing: *Towards a Theory of Educational Assessment*. London: Falmer.

Gipps, C. (1999). Socio-Cultural Aspects of Assessment. *Review of Research in Education*. Vol. 24, pp. 355-392.

Gipps, C. (2012). Beyond Testing: *Towards a Theory of Educational Assessment*. Classic Ed. London: Routledge.

Glaser, B. (1998). *Doing Grounded Theory: Issues and Discussions*. Mill Valley, California: Sociology Press.

Glaser, B. G. & Strauss, A. L. (1967). *The Discovery of Grounded Theory: strategies for qualitative research*. Hawthorne, New York: Aldine de Gruyter.

Gray, D. (2004). *Doing Research in the Real World*. London: Sage Publications.

Guest, G., MacQueen, K. M. & Namey, E. E. (2012). Applied Thematic Analysis. Thousand Oaks, California: Sage Publications.

Haertel, E. H. & Herman, J. I. (2005). A Historical Perspective on Validity Arguments for Accountability Testing. *Yearbook of the National Society for the Study of Education*. 104 (2), pp. 1-34.

Hammond, M. & Wellington, J. (2013). *Research Methods: The Key Concepts*. London: Routledge.

Han, M. & Yang, X. (2001). Educational Assessment in China: lessons from history and future prospects. *Assessment in Education: Principles, Policy & Practice*. 8 (1), pp. 5-10.

Harlen W. & Deakin Crick, R. (2002). *A systematic review of the impact of summative assessment and tests on students' motivation for learning* (EPPI-Centre Review, version 1.1). In: Research Evidence in Education Library. Issue 1. London: EPPI-Centre, Social Science Research Unit, Institute of Education.

Harlen, W. & Deakin Crick, R. D. (2003). Testing and Motivation for Learning. *Assessment in Education: Principles, Policy & Practice*. 10 (2), pp. 169-207.

Hattie, J. (1999). *Influence on Student Learning*. Inaugural Lecture: Professor of Education. University of Auckland.

Hattie, J. (2009). The Black Box of Tertiary Assessment: An Impending Revolution. In: Meyer, L. H., Davidson, S., Anderson, H., Fletcher, R., Johnston, P. M. & Rees, M. (eds.). *Tertiary Assessment & Higher Education Student Outcomes: Policy, Practice & Research*. Wellington: Ako Aotearoa. Available from: https://akoaotearoa.ac.nz/download/ng/file/group-4/n3469-the-black-box-of-tertiary-assessment---john-hattiepdf.pdf. [Accessed: 17 November 2014]

Hattie, J. A. (1987). Identifying the salient facets of a model of student learning: a synthesis of meta-analyses. *International Journal of Educational Research*. 11 (2), pp. 187-212.

Hattie, J. & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*. 77 (1), pp. 81-112.

Hau, K. & Salili, F. (1996). Motivational Effects of Teachers' Ability Versus Effort Feedback on Chinese Students' Learning. *Social Psychology of Education*. 1 (1), pp. 69-85.

Hendry, G. D., Bromberger, N. & Armstrong, S. (2011). Constructive guidance and feedback for learning: the usefulness of exemplars, marking sheets and different types of feedback in a first year law subject. *Assessment & Evaluation in Higher Education*. 36 (1), pp. 1-11.

Hennink, M., Bailey, A. & Hutter, I. (2011). *Qualitative Research Methods*. London: Sage Publications.

Higgins, R. (2000). "Be more critical!": Rethinking Assessment Feedback. *British Educational Research Association Conference*. Cardiff, September 7th-10th. Cardiff: Cardiff University. Available from: http://www.leeds.ac.uk/educol/documents/00001548.htm. [Accessed: 29 March 2015]

Higher Education Funding Council for England (2014). *Decline in global demand for English higher education* [Online]. Available from: http://www.hefce.ac.uk/news/newsarchive/2014/news86922.html. [Accessed: 17 November 2014]

Hill, M. F. (2011) 'Getting traction': enablers and barriers to implementing Assessment for Learning in secondary schools. *Assessment in Education: Principles, Policy & Practice*. 18 (4), pp. 347-364

Hodgen, J. & Marshall, B. (2005). Assessment for learning in English and mathematics: a comparison. *The Curriculum Journal*. 16 (2), pp. 153-176.

Hoepfl, M. C. (1997). Choosing Qualitative Research: A primer for Technology Education Researchers. *Journal of Technology Education*. 9 (1), pp. 47-63.

Hofstede, G. (2003). *Cultures and Organizations: Software of the Mind*. London: Profile Books.

Hounsell, D. (2007). Towards more sustainable feedback to students. In: Boud, D. & Falchikov, N. (eds.). *Rethinking Assessment in Higher Education: Learning for the longer term*. London: Routledge.

Hounsell, D., McCune, V., Hounsell, J. & Litjens, J. (2008). The quality of guidance and feedback to students. *Higher Education Research & Development*. 27 (1), pp. 55–67.

Huett, J. (2004). Email as an Educational Feedback Tool: Relative Advantages and Implementation Guidelines. *International Journal of Instructional Technology & Distance Learning* [Online]. 1 (6). Available from: http://itdl.org/Journal/Jun_04/article06.htm [Accessed: 29 March 2015]

Hwang, A., Ang, S. & Francesco, A. M. (2002). The Silent Chinese: The Influence of Face and *Kiasuism* on Student Feedback-Seeking Behaviors. *Journal of Management Education*. 26 (1), pp. 70-98.

International Monetary Fund (2014). *World Economic and Financial Surveys: World Economic Outlook Database (October 2014 Edition)* [Online]. Available from: http://www.imf.org/external/pubs/ft/weo/2014/02/weodata/index.aspx. [Accessed: 17 November 2014]

Irons, A. (2008). *Enhancing Learning through Formative Assessment and Feedback*. London: Routledge.

James, D. (2000). Making the graduate: Perspectives on student experience of assessment in higher education. In: Filer, A. (ed.). *Assessment: Social Process and Social Product*. London: Routledge/Falmer.

James, M. (1992). Assessment for Learning. *Annual Conference of the Association for Supervision and Curriculum Development, Assembly Session on Critique of*

Reforms in Assessment and Testing in Britain. New Orleans, Louisiana, April 4th –7th.

James, M. (2000). Measured lives: the rise of assessment as the engine of change in English schools. *The Curriculum Journal*. 11 (3), pp. 343-364.

Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing*. 27 (4), pp. 555-584.

Jonsson, A. (2012). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education*. 14 (1), pp. 63-76.

Jowett, A. J. (1991). The demographic responses to famine: The case of China 1958-61. *GeoJournal*. 23 (2), pp. 135-146.

Kellough, R. D. & Kellough, N. G. (1999). *Secondary School Teaching: A Guide To Methods And Resources; Planning For Competence*. Upper Saddle River, New Jersey: Prentice Hall.

Klenowski, V. (2009). Assessment for Learning Revisited: an Asia-Pacific perspective. *Assessment in Education: Principles, Policy and Practice*. 16 (3), pp. 263-268.

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), pp. 254-284.

Ko, V. (2012). Trouble Down South: Why Hong Kong and Mainland Chinese Aren't Getting Along. *Time* [Online]. Available from: http://world.time.com/2012/01/24/trouble-down-south-why-hong-kong-and-mainland-chinese-arent-getting-along. [Accessed: 17 November 2014]

Koestner, R., Zuckerman, M. & Koestner, J. (1987). Praise, involvement, and intrinsic motivation. *Journal of Personality and Social Psychology*. 53 (2), pp. 383-390.

Koh, L. C. (2008). Refocusing formative feedback to enhance learning in pre-registration nurse education. *Nurse Education in Practice*. 8 (4), pp. 223-230.

Kulhavy, R. W. (1977). Feedback in Written Instruction. *Review of Education Research*. 47 (1), pp. 211-232.

Kvale, S. & Brinkmann, S. (2009). *Interviews: Learning the Craft of Qualitative Research interviewing*. 2nd Ed. Los Angeles, California: Sage Publications.

Lapan, S. D., Quartaroli, M. T. & Riemer, F. J. (eds.) (2012). *Qualitative Research: An Introduction to Methods and Designs*. San Francisco, California: Jossey-Bass.

Leahy, S. & Wiliam, D. (2012). From teachers to schools: scaling up professional development for formative assessment. In: Gardner, J. (ed.). *Assessment and Learning*. 2nd Ed. London: Sage Publications.

Leavy, P. (2011). *Oral History: Understanding Qualitative Research*. New York: Oxford University Press.

Lee, I. (2011). Formative Assessment in EFL Writing: An Exploratory Case Study. *Changing English: Studies in Culture and Education*. 18 (1), pp. 99-111.

Leech, N. L. & Onwuegbuzie, A. J. (2009). A typology of mixed methods research designs. *International Journal of Methodology*. Quality & Quantity (43), pp. 265-275.

Levine, T. R., Park, H. S. & Kim, R. K. (2007). Some Conceptual and Theoretical Challenges for Cross-Cultural Communication Research in the 21st Century. *Journal of Intercultural Communication Research*. 36 (3), pp. 205-221.

Li, L. (2004). China's Higher Education Reform 1998-2003: A Summary. *Asia Pacific Education Review*. 5 (1), pp. 14-22.

Li, W. S. & Hui, S. K. F. (2007). Conceptions of assessment of Mainland China college lecturers: A technical paper analyzing the Chinese version of COA-III. *The Asia-Pacific Education Researcher*. 16 (2), pp. 185-198.

Lincoln, Y. S. (2009). Ethical Practices in Qualitative Research. In: Mertens, D. M. & Ginsberg, P. E. (eds.). *The Handbook of Social Research Ethics*. Thousand Oaks, California: Sage Publications.

Lincoln, Y. S. & Guba, E. G. (1985). *Naturalistic Inquiry*. Newbury Park, California: Sage Publications.

Lincoln, Y. S. & Guba, E. G. (1989). *Fourth Generation Evaluation*. Newbury Park, California: Sage Publications.

Lindsay, G. (2010). Ethical Considerations and Legal Issues in Educational Research. In: Hartas, D. (ed.). *Educational Research and Inquiry: Qualitative and Quantitative approaches*. London: Continuum Publishing.

Lipnevich, A. A. & Smith, J. K. "I really need feedback to learn:" students' perspectives on the effectiveness of the differential feedback messages. *Educational Assessment, Evaluation and Accountability*. 21 (4), pp. 347-367.

Looney, J. & Wiliam, D. (2005). England: implementing formative assessment in a high-stakes environment. In: Looney, J. (ed.). *Formative assessment: improving learning in secondary classrooms*. Paris: Organisation for Economic Cooperation and Development.

Manion, M. (1994). Survey Research in the Study of Contemporary China: Learning from Local Samples. *The China Quarterly*. No. 139, pp. 741-765.

Mansell, W., James, M. & Assessment Reform Group (2009). *Assessment in schools. Fit for purpose? A Commentary by the Teaching and Learning Research Programme*. London: Economic and Social Research Council, Teaching and Learning Research Programme.

Marshall, B. and Drummond, M. J. (2006) How teachers engage with Assessment for Learning: lessons from the Classroom. *Research Papers in Education*. 21 (2), pp. 133-149.

Mason, B. J., & Bruning, R. (2001). *Providing Feedback in Computer-based Instruction: What the Research Tells Us* [Online]. Available from: http://dwb4.unl.edu/dwb/Research/MB/MasonBruning.html. [Accessed: 17 November 2014]

Masson, M.-E. (2011). *Collecting student perceptions of feedback through interviews*. Hong Kong, August 8th – 10th. Hong Kong: The Chinese University of Hong Kong. Available from: http://www.paaljapan.org/conference2011/ProcNewest2011/pdf/oral/2H-2.pdf. [Accessed: 17 November 2014]

Mathison, S. (1988). Why Triangulate? *Educational Researcher*. 17 (2), pp. 13-17.

Matthews, B. & Ross, L. (2010). *Research Methods: A practical guide for the social sciences*. Harlow: Longman/Pearson Education.

Maxwell, J. A. (1992). Understanding and Validity in Qualitative Research. In: Huberman, A. M. & Miles, M. B. (eds.). *The Qualitative Researcher's Companion*. Thousand Oaks, California: Sage Publications.

May, T. (2011). *Social Research: Issues, Methods and Process*. 4th Ed. Maidenhead: McGraw-Hill/Open University Press.

McLeay, F. & Wesson, D. (2014). Chinese versus UK marketing students' perceptions of peer feedback and peer assessment. *The International Journal of Management Education*. 12 (2), pp. 142-150.

Mercer, J. (2007). The challenges of insider research in educational institutions: wielding a double-edged sword and resolving delicate dilemmas. *Oxford Review of Education*. 33 (1), pp. 1-17.

Miles, M. B. (1990). New methods for qualitative data collection and analysis: vignettes and pre-structured cases. *International Journal of Qualitative Studies in Education*. 3 (1), pp. 37-51.

Milmo, C. (2008). The year a new superpower is born. *The Independent* [Online]. January 1st. Available from: http://www.independent.co.uk/news/

world/asia/2008-the-year-a-new-superpower-is-born-767522.html. [Accessed: 17 November 2014]

Ministry of Education of the People's Republic of China (2014). *List of Chinese Colleges and Universities*. Available from: http://www.moe.edu.cn/publicfiles/business/htmlfiles/moe/moe_2812/201409/175222.html. [Accessed: 17 November 2014]

Mory, E. (2004). Feedback research revisited. In: Jonassen, D. H. (ed.). *Handbook of Research on Educational Communications and Technology*. 2nd Ed. Mahwah, New Jersey: Lawrence Erlbaum.

Narciss, S. & Huth, K. (2004). How to design informative tutoring feedback for multimedia learning. In: Niegemann, H. M., Brünken, R. Leutner, D. (eds.). *Instructional design for multimedia learning*. Munster, New York: Waxmann.

National Bureau of Statistics of China (2011). *Communiqué of the National Bureau of Statistics of People's Republic of China on Major Figures of the 2010 Population Census (No. 1)* [Online]. Available from: http://www.stats.gov.cn/english/NewsEvents/201104/t20110428_26449.html. [Accessed: 17 November 2014]

Nicol, D. (2010). From monologue to dialogue: improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*. 35 (5), pp. 501-517.

Nicol, D., Thomson, A. & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education*. 39 (1), pp. 102-122.

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

Norris, N. (1997). Error, bias and validity in qualitative research. *Educational Action Research*. 5 (1), pp. 172-176.

O'Leary, M. & Brooks, V. (2014). Raising the stakes: classroom observation in the further education sector in England. *Professional Development in Education*. 40 (4), pp. 530-545.

Organisation for Economic Cooperation and Development (2010). *Shanghai and Hong Kong: Two Distinct Examples of Education Reform in China* [Online]. Available from: http://www.oecd.org/dataoecd/34/45/46581016.pdf. [Accessed: 17 November 2014]

Parsons, C. (2007). Web-Based Surveys: Best Practices Based on the Research Literature. *Visitor Studies*. 10 (1), pp. 13-33.

Patton, M. Q. (2002). *Qualitative Research & Evaluation Methods*. 3rd Ed. Thousand Oaks, California: Sage Publications.

Perrenoud, P. (1998). From Formative Evaluation to a Controlled Regulation of Learning Processes. Towards a wider conceptual field. *Assessment in Education: Principles, Policy & Practice*. 5 (1), pp. 85-102.

Phuong-Mai, N., Terlouw, C. & Pilot, A. (2005). Cooperative learning vs Confucian heritage culture's collectivism: confrontation to reveal some cultural conflicts and mismatch. *Asia Europe Journal*. 3 (3), pp. 403-419.

Platt, J. (1985). Weber's Verstehen and the History of Qualitative Research: The Missing Link. *The British Journal of Sociology*. 36 (3), pp. 448-466.

Popham, W. J. (2001). Teaching to the test? *Educational Leadership*. 58 (6), pp. 16–20.

Poulos, A. & Mahony, M. J. (2008). Effectiveness of feedback: the students' perspective. *Assessment & Evaluation in Higher Education*. 33 (2), pp. 143-154.

Preissle, J. (2006). Envisioning qualitative inquiry: a view across four decades. *International Journal of Qualitative Studies in Education*. 19 (6), pp. 685-695.

Price, M., Handley, K., Millar, J. & O'Donovan, B. (2010). Feedback : all that effort, but what is the effect? *Assessment & Evaluation in Higher Education*. 35 (3), pp. 277-289.

Pryor, J. & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford Review of Education*. 34 (1), pp. 1-20.

Qu, W. & Yang, S. (2010). A Peer and Self-assessment Project Implemented in Practical Group Work. *Journal of Language Teaching and Research*. 1 (6), pp. 776-781.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*. 28 (1), pp. 4-13.

Ramsden, P. (2003). *Learning to Teach in Higher Education*. 2nd Ed. London: Routledge/Falmer.

Ronayne, M. (1999). Marking and Feedback. *Improving Schools*. 2 (42), pp. 42-43.

Rossman, G. B. & Rallis, S. F. (2003). *Learning in the Field: An Introduction to Qualitative Research*. 2nd Ed. Thousand Oaks, California: Sage Publications.

Rowntree, D. (1987). Assessing Students: How shall we know them? Revised Ed. London: Kogan Page.

Roy, A., Walters, P. G. P. & Luk, S. T. K. (2001). Chinese puzzles and paradoxes: Conducting business research in China. *Journal of Business Research*. 52 (2), pp. 203-210.

Ryan, G. W. & Bernard, H. R. (2003). Techniques to Identify Themes. *Field Methods*. 15 (1), pp. 85-109.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*. 18 (2), pp. 118-144.

Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice*. 5 (1), pp. 77-84.

Sadler, D. R. (2010). Beyond feedback: developing student capability in complex

appraisal. *Assessment and Evaluation in Higher Education*. 35 (5), pp. 535-550.

Sarantakos, S. (2005). *Social Research*. 3rd Ed. Basingstoke: Palgrave Macmillan.

Schunk, D. H. (1982). Effects of effort attributional feedback on children's perceived self-efficacy and achievement. *Journal of Educational Psychology*. 74 (4), pp. 548-556.

Schunk, D. H. (1983). Ability versus effort attributional feedback: Differential Effects on Self-Efficacy and Achievement. *Journal of Educational Psychology*. 75 (6), pp. 848-856.

Schunk, D. H. (1984). Self-efficacy perspective on achievement behavior. *Educational Psychologist*. 19 (1), pp. 48-58.

Schunk, D. H. (1985). Self-efficacy and classroom learning. *Psychology in the Schools*. 22 (2), pp. 208-223.

Schunk, D. H. (2003). Self-efficacy for reading and writing: Influence of modeling, goal setting, and self-evaluation. *Reading and Writing Quarterly*. 19 (2), pp. 159-172.

Schunk, D. H., & Rice, J. M. (1991). Learning goals and progress feedback during reading comprehension instruction. *Journal of Reading Behavior*. 23 (3), pp. 351–364.

Schunk, D. H. & Swartz, C. W. (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology*. 18 (3), pp. 337-354.

Scriven, M. (1967). The methodology of evaluation. In: Tyler, R. W., Gagné, R. M. & Scriven, M. (eds.). *Perspectives of curriculum evaluation*. Chicago, Illinois: Rand McNally.

Scriven, M. (1980). *The logic of evaluation*. Inverness, California: EdgePress.

Scriven, M. (1991). *Evaluation Thesaurus*. 4th Ed. Newbury Park, California: Sage Publications.

Seale, C. (1999). *The Quality of Qualitative Research: Introducing Qualitative Methods*. London: Sage Publications.

Seale, C. (2003). Quality in Qualitative Research. In: Lincoln, Y. S. & Denzin, N. K. (eds.). *Turning Points in Qualitative Research: Tying Knots in a Handkerchief*. Walnut Creek, California: AltaMira Press.

Shing, L. W. & King Fai, S. H. (2007). Conceptions of Assessment of Mainland China College Lecturers: A Technical Paper Analyzing the Chinese Version of COA-III. *The Asia-Pacific Education Researcher*. 16 (2), pp. 185-198.

Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*. 78 (1) pp. 153-189.

Shuy, R. W. (2002). In-person Versus Telephone Interviewing. In: Gubrium, J. F. & Holstein, J. A. (eds.). *Handbook of Interview Research: Context and Method*. Thousand Oaks, California: Sage Publications.

Simco, N. & Warin, J. (1997). Validity in Image-based Research: an elaborated illustration of the issues. *British Educational Research Journal*. 23 (5), pp. 661-672.

Smith, E. & Gorard, S. (2005). The dissemination of formative assessment: A short response to Black, Harrison, Hodgen, Marshall and Wiliam. *Research Intelligence*. 93 (6), p. 6.

Stefani, L. A. J. (1998). Assessment in Partnership with Learners. *Assessment & Evaluation in Higher Education*. 23 (4), pp. 339-350.

Stenbacka, C. (2001). Qualitative research requires quality concepts of its own. *Management Decision*. 39 (7), pp. 551-555.

Stening, B. W. & Zhang, M. Y. (2007). Methodological Challenges Confronted when Conducting Management Research in China. *International Journal of Cross Cultural Management*. 7 (1), pp. 121-142.

Stewart, W. (2010). *A decade on, academic who led AFL movement admits it 'isn't happening' in many classrooms* [Online]. Available from: http://www.tes.co.uk/article.aspx?storycode=6062338. [Accessed: 17 November 2014]

Stiggins, R. (2002). Assessment Crisis: The Absence Of Assessment FOR Learning. *Phi Delta Kappan*. 83 (10), pp. 758-765.

Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*. 87 (4), pp. 324-328.

Stobart, G. (2008). *Testing Times: The uses and abuses of assessment*. Abingdon: Routledge.

Stobart, G. (2010). *Assessment for Learning's mid-life identity crisis*. Cambridge Assessment Network, Network News Issue 31 [Online]. Available from: http://www.assessnet.org.uk/e-learning/file.php/1/Resources/newsletters/Network_News_-_Issue_31.pdf. [Accessed: 17 November 2014]

Strauss, A. & Corbin, J. (1998). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. 2nd Ed. Thousand Oaks, California: Sage Publications.

Sully de Luque, M. F. & Sommer, S. M. (2000). The impact of culture on feedback-seeking behaviour: An integrated model and propositions. The Academy of Management Review. 25 (4), pp. 829-849.

Sutton, R. (1995). *Assessment for Learning*. Salford: RS Publications.

Swaffield, S. (2009). The misrepresentation of Assessment for Learning – and the woeful waste of a wonderful opportunity. *Association for Achievement and Improvement through Assessment National Conference*. Bournemouth, 16th–18th September. Available from: http://cdn.aaia.org.uk/content/uploads/2010/07/The-Misrepresentation-of-Assessment-for-Learning1.pdf. [Accessed: 17 November 2014]

Taki, S. & Ramazani, Z. (2011). Improving Reading Skills through E-mail: The Case of Iranian EFL Students. *International Journal of Instructional Technology & Distance Learning* [Online]. 8 (4). Available from: http://www.itdl.org/Journal/Apr_11/article02.htm. [Accessed: 29 March 2015]

Tang, Y. (2013). A Case Study of Formative Assessment in a Chinese High School. *Indonesian Journal of Applied Linguistics*. 2 (2), pp. 216-225.

Taras, M. (2002). Using Assessment for Learning and Learning from Assessment. *Assessment & Evaluation in Higher Education*. 27 (6), pp. 501-510.

Taras, M. (2008). Assessment for learning: sectarian divisions of terminology and concepts. *Journal of Further and Higher Education*. 32 (4), pp. 389-397.

Taras, M. (2013). Feedback on feedback: Uncrossing wires across sectors. In: Merry, S., Price, M., Carless, D. & Taras, M. (eds.). *Reconceptualising Feedback in Higher Education: Developing Dialogue with Students*. Abingdon: Routledge.

Thomsen, D. K. & Brinkmann, S. (2009). An Interviewer's Guide to Autobiographical Memory: Ways to Elicit Concrete Experiences and to Avoid Pitfalls in Interpreting Them. *Qualitative Research in Psychology*. 6 (4), pp. 294-312.

Torrance, H. (2012). Formative assessment at the crossroads: conformative, deformative and transformative assessment. *Oxford Review of Education*. 38 (3), pp. 323-342.

Tracy, S. J. (2013). *Qualitative Research Methods: Collecting Evidence, Crafting Analysis, Communicating Impact*. Chichester: Wiley-Blackwell.

Trading Economics (2014). *China Foreign Exchange Reserves* [Online]. Available from: http://www.tradingeconomics.com/china/foreign-exchange-reserves. [Accessed: 17 November 2014]

Tweed, R. D. & Lehman, D. R. (2002). Learning considered within a cultural context: Confucian and Socratic approaches. *American Psychologist*. 57 (2), pp. 89-99.

UK Council for International Student Affairs (2014). *International student statistics: UK higher education* [Online]. Available from: http://www.ukcisa.org.uk/Info-for-universities-colleges--schools/Policy-research--statistics/Research--statistics/International-students-in-UK-HE/#Top-non-EU-sending-countries. [Accessed: 17 November 2014]

United Nations Educational, Scientific and Cultural Organisation Institute of Statistics (2011). *Country and Regional Profiles* [Online]. Available from: http://stats.uis.unesco.org/unesco/TableViewer/document.aspx?ReportId=198&IF_Language=eng [Accessed: 17 November 2014]

United Nations Educational, Scientific and Cultural Organisation Institute of Statistics (2012). *Global Education Digest 2012: Opportunities lost: The impact of grade repetition and early school leaving* [Online]. Available from: http://unesdoc.unesco.org/images/0021/002184/218449e.pdf. [Accessed: 17 November 2014]

University of Warwick (2013). *Research Code of Practice* [Online]. Available from: http://www2.warwick.ac.uk/services/rss/researchgovernance_ethics/research_code_of_practice/130318_research_code_of_practice.pdf. [Accessed: 17 November 2014]

Värlander, S. (2008). The role of students' emotions in formal feedback situations. *Teaching in Higher Education*. 13 (2), pp. 145-156.

Verma, G. K. & Mallick, K. (1999). *Researching Education: perspectives and techniques*. London: Taylor and Francis.

Walker, A. (2011). *Might China's economy stumble?* [Online]. Available from: http://www.bbc.co.uk/news/business-13802453. [Accessed: 17 November 2014]

Walker, M. (2009). An investigation into written comments on assignments: do students find them usable? *Assessment & Evaluation in Higher Education*. 34 (1), pp. 67-78.

Walliman, N. (2011). *Research Methods: The Basics*. London: Routledge.

Wang, G. (1996). Educational Assessment in China. *Assessment in Education: Principles, Policy & Practice*. 3 (1), pp. 75-88.

Wang, J. & Wu, J. (2012). Error Feedback on Students' Writings by Chinese Tertiary Teachers of English. *Chinese Journal of Applied Linguistics*. 35 (3), pp. 287-300.

Wang, P. (2010). Dealing with English Majors' Written Errors in Chinese Universities. *Journal of Language Teaching and Research*. 1 (3), pp. 194-205.

Wang, W. & Lam, A. S. L. (2009). The English Language Curriculum for Senior Secondary School in China: Its Evolution from 1949. *RELC Journal*. 40 (1), pp. 65-82.

Wang, X. (2007). A Case Study of Transition from Summative to Formative Assessment in Chinese Context. *World Congress on Engineering and Computer Science*. San Francisco, October 24th – 26th. Available from: http://www.iaeng.org/publication/WCECS2007/WCECS2007_pp478-483.pdf. [Accessed: 17 November 2014]

Wang, X. (2008). Teachers' Views on Conducting Formative Assessment in Chinese Context. *Engineering Letters*. 16 (2), pp. 231-235.

Wang, X. & Lê, T. (2006). The Discourse of Transition from Summative to Formative Assessment in China. In: *AARE Conference*. WAN06117. Adelaide, November 27th – December 1st. Available from: http://www.aare.edu.au/data/publications/2006/wan06117.pdf. [Accessed: 17 November 2014]

Wei, L. (2010). Formative Assessment: Opportunities and Challenges. *Journal of Language Teaching and Research*. 1 (6), pp. 838-841.

Wei, L. (2011). Formative Assessment in Classrooms: Operational Procedures. *Journal of Language Teaching and Research*. 2 (1), pp. 99-103.

Wei, Y. & Chen, Y. (2004). Supporting Chinese learners of English to implement self-assessment in L2 writing. Melbourne, September 13th and 14th. *Independent Learning Conference (Oceania)*. Auckland: Independent Learning Association Oceania.

Wiener, N. (1950). *The Human Use of Human Beings: Cybernetics and Society*. Boston, Massachusetts: Houghton Mifflin Co.

Wiliam, D. (1998). Enculturating learners into communities of practice: raising achievement through classroom assessment. *European Conference on Educational Research*. Ljubljana, September 17th–20th. Available from: http://www.dylanwiliam.org/Dylan_Wiliams_website/Papers_files/ECER%2098%20paper.doc. [Accessed: 17 November 2014]

Wiliam, D. (2000). Integrating formative and summative functions of assessment. *International Congress on Mathematics Education*. Tokyo/Makuhari, July 31st – August 6th. Available from: http://www.dylanwiliam.org/Dylan_Wiliams_website/Papers_files/ICME9%20WGA10%20paper.doc. [Accessed: 17 November 2014]

Wiliam, D. (2005). Keeping learning on track: Formative assessment and the regulation of learning. In: Coupland, M., Anderson, J. & Spencer, T. (eds.). *Making Mathematics Vital: Proceedings of the Twentieth Biennial Conference of the Australian Association of Mathematics Teachers*. Adelaide: Australian Association of Mathematics Teachers Inc. Available from: www.aamt.edu.au/content/download/19063/252036/file/mm-vital.pdf. [Accessed: 17 November 2014]

Wiliam, D. (2006a). Assessment for Learning: why, what and how. In: *Excellence in Assessment: Assessment for Learning*. A supplement to the Cambridge Assessment Network Assessment for Learning Seminar. Cambridge, November 15th. Available from: http://www.assessnet.org.uk/e-learning/file.php/1/Resources/Excellence_in_Assessment/Excellence_in_Assessment_-_Issue_1.pdf. [Accessed: 17 November 2014]

Wiliam, D. (2006b). Formative assessment: getting the focus right. *Educational Assessment*. 11 (3 & 4), pp. 283-289.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*. 37 (1), pp. 3-14.

Wiliam, D. (2012). Feedback: Part of a System. *Educational Leadership*. 70 (1), pp. 30-34.

Wiliam, D. & Black, P. (1996). Meanings and Consequences: a basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal*. 22 (5), pp. 537-548.

Winter, G. (2000). A Comparative Discussion of the Notion of 'Validity' in Qualitative and Quantitative Research. *The Qualitative Report* [Online]. 4 (3 & 4). Available from: http://www.nova.edu/ssss/QR/QR4-3/winter.html. [Accessed: 17 November 2014]

Xiao, Y. & Carless, D. R. (2013). Illustrating Students' Perceptions of English Language Assessment: Voices from China. *RELC Journal*. 0 (0), pp. 1-21.

Yang, M. & Carless, D. (2013). The feedback triangle and the enhancement of dialogic feedback processes. *Teaching in Higher Education*. 18 (3), pp. 285-297.

Yin, R. K. (2011). *Qualitative Research from Start to Finish*. New York: The Guildford Press.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*. 45 (4), pp. 477-501.

Yu, F. Y. & Yu, H. J. J. (2002). Incorporating e-mail into the learning process: its impact on student academic achievement and attitudes. *Computers & Education*. 38, pp. 117-126.

Yu, L. & Suen, H. K. (2005). Historical and Contemporary Exam-driven Education Fever in China. *KEDI Journal of Educational Policy*. 2 (1), pp. 17-33.

Zhan, Y. & Wan, Z. (2010). Perspectives on the Cultural Appropriateness of Assessment for Learning in the Chinese Context. *Educate*. 10 (2), pp. 9-16.

Zhao, H. (2010). Investigating learners' use and understanding of peer and teacher feedback on writing: A comparative study in a Chinese English writing classroom. *Assessing Writing*. 15 (1), pp. 3-17.

Zhu, C. (2012). Providing formative feedback to students via emails and feedback strategies based on student metacognition. *Reflecting Education*. 8 (1), pp. 78-93.

Appendices

Appendix 1: Covering Letter

Student Questionnaire

Doctoral Study on Teacher Feedback in China

Dear Student,

As you may already know, my name is Chris Cookson and I am an English teacher here at [*redacted*]. I am also a doctoral student (博士候选人) at the University of Warwick (华威大学) in England and I am doing my research here. My study and the questionnaire are about the feedback (反应) you have received from your Chinese teachers on your assessments while at [*redacted*].

Before you complete this questionnaire, I will explain some of the words I have just used. First, an assessment can be many things: an exam, a test or a quiz. It can also be a presentation, a project or another kind of written or oral assignment. An assessment measures learning in some way. The feedback you receive after you complete an assessment could be a mark/grade, oral or written comments, or it could be a combination of these.

The questionnaire should take around 10 minutes to answer. Please choose only one option per question and write your answers in English. Please do not worry about grammar mistakes! Just try your best to answer the questions as accurately as possible. I appreciate your participation very much.

The separate document – "Participant Rights" – provides a full explanation of how I will protect your identity and your rights. After you have completed the questionnaire, please return it to me by e-mail to: [*redacted*]. If you would like to remain anonymous (匿名), you may print out the document and put in my mailbox labelled "Chris Cookson" which is located on the first floor of 文南楼, 外国语言文学.

If you have any questions or concerns, you can e-mail me at the same e-mail address.

Thank you and I look forward to receiving your completed questionnaire!

Best wishes,

Chris Cookson

Appendix 2: Participant Rights Letter

Participant Rights

Dear Student,

Thank you for your interest in taking part in my study. Your opinions and ideas are valuable and of great interest to me.

By filling in and returning the questionnaire to me, you give me permission to include this information in my educational doctorate thesis (博士论文) and any future publications (刊物). My research meets the ethical requirements (道德要求) of the British Educational Research Association (英国教育研究协会). This means that I will keep any information you share with me anonymous (匿名). This includes not mentioning your name or the name of the university. In addition, I will not share your contact details with anyone else.

I will save all information you share with me in a place on my computer that is protected with a password. I would be grateful if you answered as many questions as possible. You may, however, miss any questions which you prefer not to answer. You may also withdraw (退出) from the study or revoke (撤消) any information you have shared with me at any time.

If you would like me to share with you the results of my study, I would be happy to do this; please state this in your e-mail when you send me the completed questionnaire.

Please feel free to contact me if you have any questions or concerns.

Thank you once again for your willingness to participate. I look forward to receiving your completed questionnaire.

Best wishes,

Chris Cookson

Appendix 3: Student Questionnaire

Institute of EDUCATION

Student Questionnaire

Note: Please remember to consider only the feedback you have received from your Chinese teachers (中国籍教师) here at [*redacted*].

Question 1: How effective are the following types of feedback for helping you to improve your overall learning?
Please choose one box for each type of feedback by placing a "√" inside it.

Feedback type	Very effective	Effective	Not very effective	Not at all effective	Not sure
Marks/grades (on any kind of assessment)					
Written comments that tell you what you have done well/correctly					
Written comments that tell you what you have done poorly/incorrectly					
Written comments that tell you what you need to do differently in the future					
Written comments that explain how to make improvements					
Comments written together at the end of a written assignment					
Annotations (批注) in the body or margin (页边的空白) of a written assignment					
Oral comments given collectively to groups or the whole class during lessons					
Oral comments given to you directly about your work during/after lessons					

Question 2: How often do you receive each type of feedback on your assessments?
Please choose one box for each type of feedback by placing a "√" inside it.

Feedback type	Always	Often	Sometimes	Seldom	Never
Marks/grades (on any kind of assessment)					
Written comments that tell you what you have done well/correctly					
Written comments that tell you what you have done poorly/incorrectly					
Written comments that tell you what you need to do differently in the future					
Written comments that explain how to make improvements					
Comments written together at the end of a written assignment					
Annotations (批注) in the body or margin (页边的空白) of a written assignment					
Oral comments given collectively to groups or the whole class during lessons					
Oral comments given to you directly about your work during/after lessons					

Question 3: What is the overall effect of the feedback you receive on the following aspects?
Please choose one box for each aspect by placing a "√" inside it.

Aspect	Very positive	Positive	Negative	Very negative	No effect/ Not sure
Your ability to achieve better results on your assessments					
Your ability to improve your studying habits*					
Your motivation**					
Your self-confidence***					

*studying habits = the way(s) in which you study and prepare for your classes or assessments
**motivation (动力) = your interest, desire or enthusiasm for achieving a goal
***self-confidence (自信) = your belief in your ability to achieve a task or a goal

Question 4: When do you actually usually receive feedback on your assessments?
Please choose one box on the right by placing a "√" inside it.

While you are working on the assessment	
After you have received the assessment back	
Both	

Question 5: When do you prefer to receive feedback on your assessments?
Please choose one box on the right by placing a "√" inside it.

While you are working on the assessment	
After you have received the assessment back	
Both	

Question 6: Please complete the sentence below:

In my opinion, the most important purpose of feedback is ...

Question 7: Please complete the sentence below:

Feedback is most helpful for me when …

Question 8: Would you like to offer any additional comments about feedback?

Your comments:

Question 9: What is your gender? Please choose one box by placing a "√" inside it.

Male	Female

Question 10: Which type of student are you? Please choose one box by placing a "√" inside it.

Undergraduate (本科生)	
Master (硕士生)	
Doctoral/research (博士生)	

Question 11: Would you be willing to take part in a short interview so that I may find out more about your opinions and ideas on feedback? If yes, please provide your details so that I may contact you:

Name	E-mail	Phone number

That is the end of the questionnaire.
Thank you very much for your time and participation!

Appendix 4: Generic Interview Schedule

I. Opening

A. (**Establishing Rapport**) Hello. Thank you very much for agreeing to this interview. I appreciate your time.

B. (**Purpose**) As you know, I'm carrying out research into the feedback students here at this university receive from their Chinese teachers. I found the answers you provided in the questionnaire very interesting and I would like to ask you some further questions.

C. (**Motivation**) The information you give me will help me to better understand how Chinese students at this university experience feedback and the effect it has on them.

D. (**Rights**) Before we begin, I would just like to let you know that you may skip any questions you feel uncomfortable answering and we can pause or stop the interview at any time. If you don't understand any of the questions, please let me know and I will try to explain them more clearly. May I have your permission to record the interview? The recording and any information you share with me will be kept safe and confidential.

E. (**Time Line**) The interview should take about 20 minutes. Do you have any questions before we begin? Are you ready to begin answering some questions?

[Showing a copy of the questionnaire completed by the interviewee]

[Invitation to the interviewee to produce any samples of written feedback they may have brought with them for illustrative and evocative purposes – these will be called upon throughout the interview]

[Transition to the first topic]

II. Main Body

A. Interviewee's area of study

1. Could you please tell me your subject and in which department you are studying it?

[Transition to the next topic]

B. Effectiveness of different types of feedback

1. According to your answers to Question 1, [feedback type(s)] is/are most effective in helping you to improve your overall learning. Could you please tell me why?
2. You also stated that [feedback type(s)] is/are least effective in helping you to improve your overall learning. Could you please tell me why?

[Transition to the next topic]

C. Frequency of feedback
    1. According to your answers to Question 2, you [frequency] receive [type(s) of feedback]. Why do you believe that is?
    2. You also stated that you [frequency] receive [type(s) of feedback]. Why do you believe that is?

[Transition to the next topic]

D. Psychological effect of feedback
    1. In your answers to Question 3, you stated that the feedback you receive generally has a [effect and effect magnitude] on [aspect(s)]. Could you please give me some examples or describe the effects in more detail?
    2. You also stated that the feedback you receive generally has a [effect and effect magnitude] on [aspect(s)]. Could you please also give me some examples or describe the effects in more detail?

[Transition to the next topic]

E. Timing of feedback
    1. In your answer to Question 4, you wrote that you usually receive feedback [timing category]. Why do you believe that is?
    2. In your answer to Question 5, you wrote that you prefer to receive feedback [timing category]. Could you please tell me why?

[Transition to the next topic]

F. Purpose of feedback
    1. In Question 6, I asked you what you believe the most important purpose of feedback is. You wrote: [quoted/paraphrased response]. [tailored question

eliciting the interviewee's views on the purpose(s) of feedback in greater depth].

2. To what/whom does the feedback you receive most frequently compare you to? For example, to a set of criteria, other students, yourself?
3. To what/whom do you believe feedback should compare you to? Why?

[Transition to the next topic]

G. Characterisation of effective feedback
    1. In Question 7, I asked you to describe feedback when it is most helpful. You wrote: [quoted/paraphrased response]. [tailored question eliciting the interviewee's views on the purpose(s) of feedback in greater depth].
    2. Do you find praising feedback helpful? Why/why not?
    3. Does most of the praising feedback you receive refer to the assessment or to yourself personally?
    4. Do you find critical feedback helpful? Why/why not?
    5. Does most of the critical feedback you receive refer to the assessment or to yourself personally?
    6. Is there anything else that helpful feedback should or shouldn't include?

[Transition to the last topic]

H. Additional comments
    1. In Question 8, you offered some additional comments. You wrote: [quoted/paraphrased response]. [tailored question based on the interviewee's questionnaire response].

III. Closing

A. (**Maintaining Rapport**) Thank you. Those are all the questions I have. I appreciate the time you have taken to be interviewed.
B. (**Additional comments**) Before we conclude the interview, are there any final comments you would like to add?
C. (**Follow-up**) If I have any questions regarding any of the answers you have given me today, will you allow me to contact you? Thank you again. Goodbye.

Appendix 5: Ethical Approval Form

THE UNIVERSITY OF WARWICK
Institute of
EDUCATION

Application for Ethical Approval for Research Degrees (MA by research, MPHIL/PhD, EdD)

Name of student: Chris Cookson

MA By research	**EdD**	PhD

Project title: **An Investigation into Chinese Students' Experiences of and Attitudes Towards Tutor Feedback at a Chinese University**

Supervisor: **Dr. Val Brooks**

Funding Body (if relevant): **N/A**

Please ensure you have read the Guidance for the Ethical Conduct of Research available in the handbook.

Methodology
Please outline the methodology, e.g. observation, individual interviews, focus groups, group testing etc.

Data will be collected in two phases and through two mechanisms, the first of which is a self-administered questionnaire (which will be made available for download online through the university's own website) and the second a semi-structured interview (which will be carried out in person).

Participants
Please specify all participants in the research including ages of children and young people where appropriate. Also specify if any participants are vulnerable, e.g. children; as a result of learning disability.

Primarily Chinese undergraduate, but also postgraduate students, at least 18 years of age. A small number of Chinese teachers may also be included.

Respect for participants' rights and dignity
How will the fundamental rights and dignity of participants be respected, e.g. confidentiality, respect of cultural and religious values?

I have spent 26 years living, working and studying outside of the UK in, in some cases, very different cultures on 4 different continents. I am fully aware of not only the need for sensitivity toward and respect of cultural and religious values in general terms, but, after having lived and worked in China for over 5 years, I am

also aware of those specific to this country and its citizens. In addition, I have taught for over 6 years, during which time I have developed a greater understanding of the psychology or internal dynamics of students when communicating with teachers. I do not employ specific mechanisms in protecting people's socio-cultural values other than to inform myself beforehand as to what these may be and to always be mindful of what to say or avoid saying. Anything else would be a matter of common sense.

Please see the next section for how participants' confidentiality will be protected.

Privacy and confidentiality

How will confidentiality be assured? Please address all aspects of research including protection of data records, thesis, reports/papers that might arise from the study.

As a general rule, confidentiality, while certainly a worthy goal, cannot be guaranteed. This arises out of the necessity of reporting the findings of the study in my thesis. However, anonymity and the greatest possible degree of non-traceability for the participants and the host institution will be ensured through the use of pseudonyms, code names and general descriptions (e.g. "an institution of higher learning in China").

I will store all electronic information I receive from participants in a password-protected location on my computer which itself is also protected by a password. Any hard data that I may collect will be kept in a locked container.

Consent - will prior informed consent be obtained?

- from participants? **Yes**/No from others? **Yes**/No

- explain how this will be obtained. If prior informed consent is not to be obtained, give reason:

In order to meet the requirements of prior informed consent, participants (who need to have sufficient reasoning faculties) need to be made aware of the background, implications and all other relevant details of the study before providing any information through any data collection method. I will render these requirements through a covering letter giving this information and a separate document detailing their rights. I will reiterate their rights before conducting interviews. In the case of the university (the "gatekeeper"), this has already been rendered verbally and explicitly by several senior departmental staff members.

- will participants be explicitly informed of the student's status?

Yes.

Competence

How will you ensure that all methods used are undertaken with the necessary competence?

Competence in research methodology is a skill that is acquired continuously and incrementally through knowledge of the relevant theory and through experience. Through my readings and module assignments completed in the first year of my EdD, I gained a good overview of the theory. I also gained – albeit limited – experience conducting research with live participants through my FRM study. In addition, I have been given much guidance in this area by my supervisor. If I am ever uncertain as to how to deal with a particular research situation, I will – as I have done on several occasions already – confer with her.

Protection of participants

How will participants' safety and well-being be safeguarded?

The research topic is largely independent of socio-political themes and the research questions are not of intrusive or personal nature. Therefore, there exists little to no danger of any negative emotional effects. The questionnaire will be filled in by the participants in the physical absence of the researcher. The interviews will be conducted in an environment which is familiar to the participants and visible to or accessible by other students, teachers and staff (e.g. offices, meetings rooms, classrooms, etc.). Therefore, participants will not perceive any potential for physical harm (through the researcher) to exist. Participants will be told that they may skip any questions – both on the questionnaire and during the interview – which they prefer not to answer, and they will also be made aware of their right to withdraw from the study at any time. They will also be given the right to decline the interview being recorded. In addition, I am an employee of the institution where the participants are students. It is, therefore, all the more in my interest to take the question of participants' safety and well-being seriously!

Child protection

Will a DBS (Disclosure and Barring Service formerly CRB) check be needed?

Yes/No (If yes, please attach a copy.)

N/A

Addressing dilemmas

Even well planned research can produce ethical dilemmas. How will you address any ethical dilemmas that may arise in your research?

I will first refer to the Guidance for the Ethical Conduct of Research before then, if necessary, consulting the ethical requirements of the British Educational Research Association (BERA). If these cannot provide me with a solution, I will direct and elucidate any ethical dilemmas I encounter to my supervisor.

Misuse of research

How will you seek to ensure that the research and the evidence resulting from it are not misused?

I have, through the various explications I have offered hitherto in this form, attested to ensuring that all data collected from participants will be used appropriately. Only third parties could violate this principle, but this will be avoided through the protection of the data by the means mentioned above. It is the responsibility of anyone citing any of my future published work containing any of the data collected through my study to do so correctly.

Support for research participants

What action is proposed if sensitive issues are raised or a participant becomes upset?

If a participant raises any sensitive issues, these will be addressed in an appropriate manner and with due professionalism. If a participant becomes upset and either requests the discontinuation of the interview outright or displays signs of wanting to do so, then I will, of course, consent to this request.

Integrity

How will you ensure that your research and its reporting are honest, fair and respectful to others?

Again, through the assurances I have made above, I am bound to honesty and fairness. As both a research student and a teacher, I am obligated to conduct myself in a fashion that is appropriate to those roles, and I will abide by this principle with the greatest earnestness.

If the meaning of a contribution made by a participant – whether in written or verbal form – is not readily apparent, I will communicate with them and request confirmation of its accuracy.

What agreement has been made for the attribution of authorship by yourself and your supervisor(s) of any reports or publications?

No such agreements have yet been made. However, in the one contribution that I have had published so far (on the website of The Association for Educational Assessment – Europe), I stated both my supervisor's name and WIE. I pledge to continue doing so if so required or appropriate until the completion, submission and final acceptance of my thesis.

Other issues?

Please specify other issues not discussed above, if any, and how you will address them.

I do not foresee any additional issues at this point.

Signed,

Research student:	Date:
Chris Cookson	May 23, 2013
Supervisor:	Date: 23/05/2013
Val Brooks	

Action

Please submit to the Research Office (Louisa Hopkins, room WE132)

Action taken

- ☑ Approved but see below
- ☐ Approved with modification or conditions – see below
- ☐ Action deferred. Please supply additional information or clarification – see below

Name Geoff Lindsay Date 29/5/13

Signature

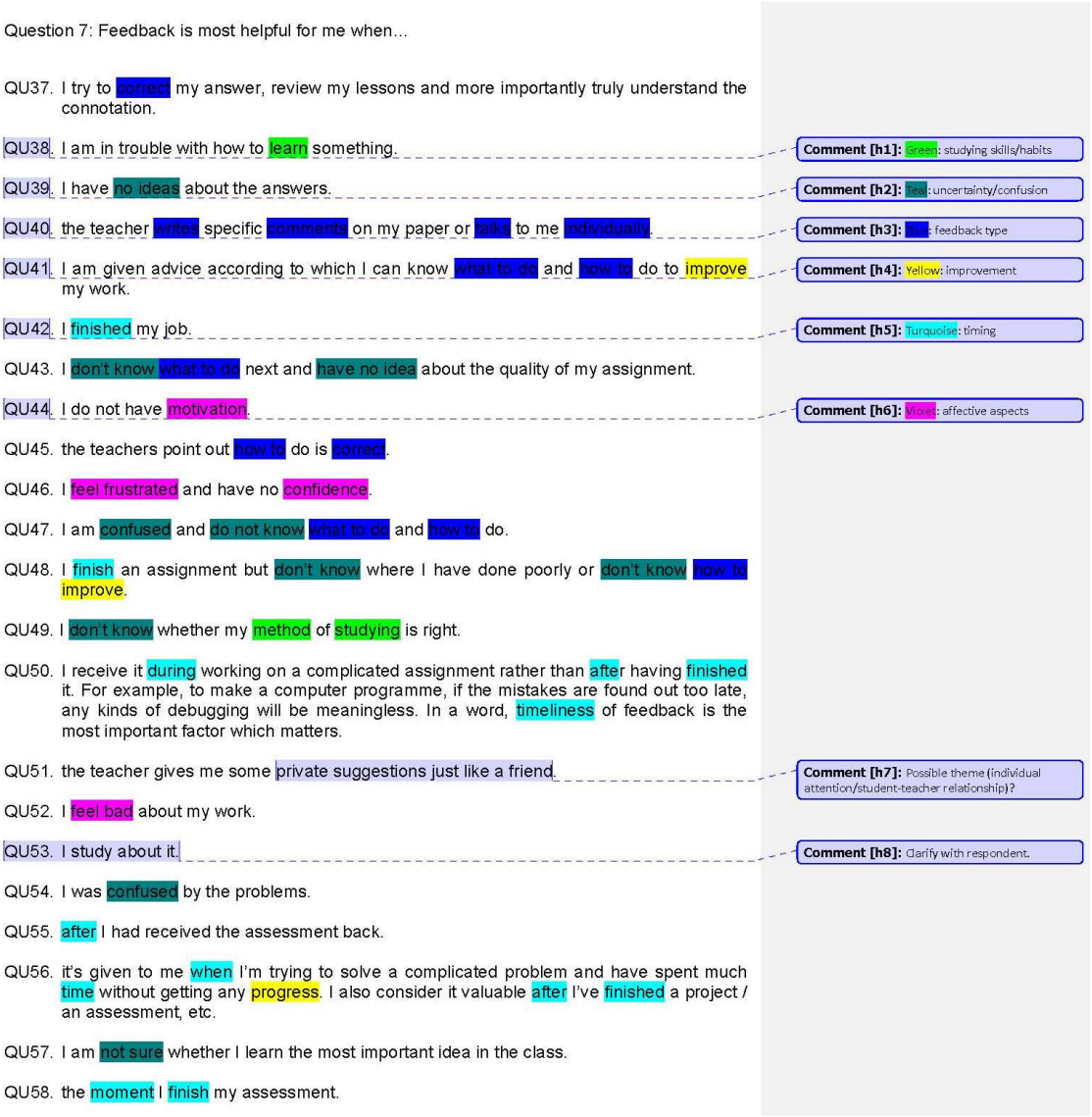
Stamped

Notes of Action A well presented, thoughtful proposal

(1) Yes, but you still have a responsibility to take action to limit risk. For example, if you produce material for the press on the research, you have a responsibility to try to explain & check the reporter understands – i.e. 'reasonable action'

(2) We agree on the basis that the thesis is ultimately yours but publications from the thesis (eg papers, conference papers) ~~should~~ will reflect also your supervisors input & joint authorship may often be appropriate

Appendix 6: Extract of Colour-Coded Questionnaire Data



Appendix 7: Extract of Transcribed Interview Data

Question:	**According to your answers to Question 1, all types of feedback are helpful in improving your overall learning. You do not seem to have a preference. Could you please tell me why?**
	QIU18
Student:	I think because, ever since I came to college, I received much less feedback than before, so I thought even a little bit of feedback would be very useful. And…the fact that the courses I take…usually the whole class contains over 200 students and it is impossible for the teacher to talk individually and all you do is submit your homework. The teacher won't mark himself…he will leave it to the assistant teacher. And…the office hours are not very available for everyone, so I seldom get the chance to see the teacher after class.
Researcher:	So, is it because you receive so little feedback that you value any you can get?
Student:	Yes.
	IU81
Student:	There are so many kinds of motivations that push me to study. But I think marks and grades rank first.
Researcher:	So, are marks and grades most effective in helping you to improve your learning?
Student:	I think it's more appropriate to say that they motivate me to study, not *help* me to study. I don't know what it's like abroad, but in China, this goes first…in the Chinese education system. Because every evaluation is based on marks and grades.
Researcher:	But you do find all kinds of feedback helpful for your learning, correct?
Student:	Yes.
Researcher:	Is there a central reason for that?
Student:	I think it's just because of my personality. Umm…I really care about how others will look upon me…how I am in others' opinions. It's nice to know they care about me and my work.
Researcher:	So, is any feedback helpful for you?
Student:	I don't like to call it helpful; just a kind of motivation, as I have said before.
	QIU30
Student:	I don't prefer a specific one because every feedback has its advantages and disadvantages. For example, grades can be very clear, but they don't further instruct you. But written feedback can give some further instruction.
Researcher:	So, what are some disadvantages of written feedback?
Student:	Sometimes…umm…written feedback is not clear. Maybe the teacher says this part of your essay is not clear and it's not well supported, and if I do not get a further office hour with the teacher, the point seems not so clear to me.

Comment [h1]: Due to receiving far less feedback than before entering university, the student has gained a heightened appreciation for that which he does receive.

Comment [h2]: The student likens any feedback she receives as the assessor taking an interest in her, which acts to motivate her due to the value she places in others' opinions.

Comment [h3]: The student sees the value in any type of feedback and feels that the weaknesses inherent in one type can be offset by its strengths.