THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

The Contexts of Assessment in EFL Classrooms
in Two High Schools in Vietnam

Thanh Nhan Le

BA (Quy Nhon Teacher Training College), MA in English Linguistics (College of Foreign
Languages, Vietnam National University - Hanoi)

A thesis submitted for the degree of Doctor of Philosophy at
The University of Queensland in 2015

School of Education

Abstract

The aim of this project was to investigate the contexts of assessment in EFL classrooms in high schools in Vietnam. While in many countries around the world school administrators and teachers have to administer high-stakes tests for accountability and managerial purposes, they are often also expected to implement classroom assessment for teaching and learning improvement. EFL teachers in high schools in Vietnam also face the challenges of negotiating between these different purposes of assessment.

Using a case study approach, this study investigated the contexts of assessment in six Grade 10 to 12 EFL classrooms in two high schools in a city in central Vietnam. The research questions were: (1) What are the perceptions of the contexts of assessment held by the principals, the Grade 10 to 12 EFL teachers, and students in two high schools in Vietnam? (2) How do the Grade 10 to 12 EFL teachers implement assessment in their classrooms? (3) How do the principals, the Grade 10 to 12 EFL teachers, and students negotiate the multiple purposes of assessment in their schools and classrooms?

Policy documents, semi-structured interviews with the principals and the EFL teachers, 12 focus group interviews with 72 students, and 42 classroom observations were analysed inductively using content analysis. Delandshere's (2001) conceptual framework was used to examine the interviews and the classroom observations from four dimensions: technological, philosophical, sociological, and ethical.

The key findings of the study were first that assessment in these contexts was neither typical of a culture of testing nor typical of a culture of assessment, although elements of both existed. The principals and teachers perceived that assessment meant using tests to measure students' knowledge and assign marks, which were used for reporting and making teachers and students accountable. The students perceived that assessment was about obtaining marks which were used to rank, monitor, and control them. Because marks were high-stakes, the principals and teachers were concerned about accuracy, objectivity, and fairness. The principals expected their teachers to comply with the regulations and procedures for assessment, thus limiting the teachers' autonomy. Elements of a culture of assessment were illustrated by the participants' beliefs that assessment informed teachers about their teaching and their students' learning. The teachers embedded assessment in their instruction and the students expected that they would receive quality feedback to improve their learning and achievement.

Secondly, the teachers relied on the exercises in the textbooks for assessment. These exercises mainly assessed lower-order thinking skills, especially recall of factual information,

vocabulary, and grammar rules. The teachers' feedback focused on direct corrective feedback and praise. Some teachers used marks as rewards and punishment.

The third key finding was that the participants paid more attention to the elements of a culture of testing. The principals were concerned about accuracy, objectivity, and fairness. Therefore, they expected the teachers to comply with the regulations and procedures, even though they perceived that the teachers' compliance restricted their autonomy in using assessment for teaching and learning improvement. In addition, the teachers constructed tests that helped their students obtain high marks. Although the students indicated that they wanted feedback that would make a difference to their achievement, they were very concerned about marks, and some cheated when they were tested.

This study contributes to knowledge about high school principals', teachers', and students' perceptions of assessment in Vietnam – a country where a culture of testing is still dominant, even though policy-makers have introduced a culture of assessment with the aim of improving teaching and learning. Additionally, the study contributes to an understanding of the assessment practices that Grade 10 to 12 EFL teachers implement in their classrooms. The study's findings also suggest that where assessment is high-stakes, principals, teachers, and students pay attention to the elements of a culture of testing at the expense of using assessment for teaching and learning improvement.

The limitations of this study were acknowledged. First, interviews following up the classroom observations were not conducted. Such interviews would have provided information about the reasons for the teachers' assessment practices and why they implemented them in the way they did. Second, the findings cannot be generalized to other schools in Vietnam. Third, inter-reliability checks were not undertaken. Fourth, the number of the observed lessons was small.

Based on the findings from this study, implications for future research, policy-makers, and practitioners were presented.

This study suggested that in contexts where assessment results were used for and perceived to have high-stakes consequences for students, teachers, and parents, it was difficult to develop a culture of assessment. The study indicated that assessment reform in Vietnam's high school education in general, and English instruction in particular, has not been very effective to date because major obstacles related to both perceptions and practices have not been tackled and removed. Reform of assessment in Vietnam has led to some changes in the methods of assessment rather than changes in the purposes of assessment. If the purposes of assessment are not changed and EFL teachers are not persuaded to change and provided with knowledge and skills in assessment for learning, it continues to be secondary to assessment of learning, no matter how much this is called for by policy-makers.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

No publications.

Publications included in this thesis

No publications included.

Contributions by others to the thesis

No contributions by others.

Statement of parts of the thesis submitted to qualify for the award of another degree

None.

<u>Acknowledgements</u>

During my academic life up to now, I have received a lot of encouragement, help, guidance, and expertise from many people. I would like to take this opportunity to express my gratitude to these people.

First, I would like to show my deepest gratitude to my academic supervisors, Associate Professor Dr. Christa E. van Kraayenoord (my principal supervisor) and Associate Professor Dr. Karen B. Moni (my associate supervisor). Since my first email to Dr. Kraayenoord on December 25, 2009 up to this point in time, she has spent so much time and energy assisting me with my study. Dr. Kraayenoord has seen so many ups and downs in my academic and personal life during my candidature, and she was always there giving me the right things, from encouragement, sympathy, and support to challenges and pressure. I have been lucky to be Dr. Moni's PhD student. She always knew not only what was wrong but also why it was wrong and how it could be improved. Dr. Kraayenoord and Dr. Moni, I owe you a lot, and I will pay you back by treating my future students as kindly and considerately as you both have treated me.

During my PhD candidature, the members of the Panel, Associate Professor Dr. Annemaree Carroll (Chair) and Dr. Kerryn McCluskey have given me enlightening comments and suggestions that helped me improve my knowledge about my research topic and skills in research. I would like to take this opportunity to express my deep gratitude to Dr. Carroll and Dr. McCluskey for their invaluable support.

I would like to extend my gratitude to Australian Government and Australian taxpayers for giving me an opportunity to study in Australia. My thanks are also extended to the officers at the School of Education, UQ International, UQ Student Service, and ADS offices in Australia and Vietnam.

I am grateful to the Principals, EFL teachers, and students for their participation in this study. I would like to thank the Head of the DOET of the province where I collected data for this study. I would like to thank my friends, Mr. Nguyen Quang Lau and Mr. Pham Manh Cuong for their help with my data collection.

I would like to thank Associate Professor Dr. Nguyen Hong Anh – Rector of Quy Nhon University, Dr. Ha Thanh Hai – Dean of the Department of Foreign Languages at Quy Nhon University, and my colleagues at the Department of Foreign Languages at Quy Nhon University for their support.

I would like to express my gratitude to my village friends and neighbours who were always ready to stop in the middle of their dinners to escort me on my nightly walks to my home when I was a primary school student.

Keywords

assessment, assessment practice, classroom, context, EFL teacher, high school, perception, principal, test, student

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRA code: 130303, Education Assessment and Evaluation, 90%

ANZSRA code: 130207, LOTE, ESL and TESOL Curriculum and Pedagogy (excl. Maori), 10%

Fields of Research (FoR) Classification

FoR code: 1303, Specialist Studies in Education, 90%

FoR code: 1302, Curriculum and Pedagogy, 10%

Table of Contents

List of Figures

List of Abbreviations

CLT: Communicative Language Teaching

DOET: Department of Education and Training (provincial level)

EFL: English as a foreign language

FL: Foreign language

GEDE: General Education Diploma Examination

GTM: Grammar Translation Method

L: Listening

L2: Second language

LF: Language focus

MOET: Ministry of Education and Training (central level)

NUEE: National University Entrance Examination

Ob: Observation

PA: Principal in School A

PB: Principal in School B

R: Reading

S: Speaking

S: Student (S1, S2, …, S12: 12 students from each class)

T: Testing

TA: Teacher in School A (TA1, TA2, TA3: 3 teachers in School A)

TB: Teacher in School B (TB1, TB2, TB3: 3 teachers in School B)

TR: Test return

W: Writing

Chapter 1. Introduction

This chapter introduces the study, the context in which the study was conducted, the statement of the problem, the aim of the study, and the research questions. The chapter ends with a discussion of the significance of the study.

1.1. Introduction

Assessment is an important part of teaching and learning (Chappuis, Stiggins, Chappuis, & Arter, 2012; Harris & Brown, 2008) because it may positively or negatively influence these areas (Chan, 2007). When integrated into instruction for formative purposes (i.e., "to promote, not merely to judge or grade, student success") (Stiggins, 2005, p. 326), assessment helps improve the quality of teaching and learning (Barjesteh & Niknezhad, 2013; Black & Wiliam, 1998a, 1998b; Rohrer & Pashler, 2010; Tierney, 2014). In contrast, when too much emphasis is put on the use of assessment results as a measure to make students, teachers, and schools accountable for their learning and teaching, assessment may "prevent and drive out thoughtful classroom practices" (Shepard, 2000, p. 9). In response to high-stakes testing (i.e., testing that has appreciable consequences on individuals, schools or educational systems (G. T. L. Brown & Harris, 2009; Solomon, 2002), students may be challenged by emotional pressures (Weeden, Winter, & Broadfoot, 2002), become demotivated, anxious, or pursue surface learning (Harlen, 2008), and spend too much time preparing for tests and examinations (M. G. Brooks & J. G. Brooks, 1999; R. Cohen, 1990; Stobart, 2008). Teachers may narrow the curriculum (Gipps, 1994; Harlen, 2007) and teach to the test rather than try to improve student learning (G. T. L. Brown, Lake, & Matters, 2009; Choi, 2008). Teachers' beliefs related to whether assessment helps improve learning and teaching; whether they think that assessment makes themselves, their students and schools accountable; and whether they spend too much time teaching to the tests are strongly influenced by the purposes and practices of assessment that are prioritised in their contexts (Harris & Brown, 2009; Pryor & Lubisi, 2002).

Contexts of assessment have been characterised as a culture of assessment or a culture of testing (Estaji, 2011; Inbar-Lourie, 2008a; Kleinsasser, 1995; Lynch, 2001). A culture of assessment and a culture of testing are differentiated according to their purposes, the people involved in the assessment process, types of assessment, time and frequency of assessment, and support and aids permitted in the assessment processes (Kleinsasser, 1995; Wolf, Bixby, Glenn, & Gardner, 1991). In a culture of testing, tests and examinations are used to generate scores for administrative and accountability purposes (Kleinsasser, 1995; Stiggins, 2002; Wolf et al., 1991). In

a culture of assessment, alternative assessments such as observations, oral presentations, exhibitions, self-assessment, peer-assessment, and portfolios (Klenowski, 2009; McMillan, 2001; Watt, 2005; Wolf et al., 1991) may be used to obtain evidence about student learning in order to modify teaching and improve learning (Black & Wiliam, 1998a; Klenowski, 2009; Lynch, 2001). (Detailed discussions of the characteristics of a culture of testing and a culture of assessment are provided in Chapter 2).

Although a culture of testing and a culture of assessment have distinct characteristics, they should not be seen as two rival cultures. Inbar-Lourie (2008a) argues that a culture of testing is not replaced by the introduction of a culture of assessment. Rather, they coexist and complement each other (Nagy, 2000).

In reality, teachers may face a tension between two opposing demands in assessing their students (Blackmore, 1988; Pryor & Lubisi, 2002). On the one hand, they have to administer high-stakes tests and examinations to generate scores to report to parents; to make students accountable for their learning; to show their accountability; and to rank, classify, certify, or select students (Harlen, 2008). On the other hand, they want to, and are encouraged to, implement classroom assessment for the purposes of modifying teaching and improving learning (Berry, 2011a; Hargreaves, 2005; Remesal, 2007; Tierney, 2014).

This tension has been found in research undertaken in some contexts. For example, in a study of the assessment practices of two South African teachers in Grades 4 to 6, the teachers used a lot of formative assessments, but they worried that they had no concrete evidence to show parents and administrators that they had assessed their students (Pryor & Lubisi, 2002). These teachers' concern may derive from the emphasis on accountability and on reporting learners' achievement of specific outcomes in the South African education system (Vandeyar & Killen, 2007).

In the context of Hong Kong, where high-stakes tests and examinations have existed for a long time (Berry, 2011a; Lee & Coniam, 2013), Brown, Kennedy, Fok, Chan, and Yu (2009) investigated the perceptions of assessment held by 288 teachers from 12 primary and three secondary schools. The study showed that the teachers strongly believed that assessment helped improve student learning but they also strongly supported the notion that assessment made students accountable for their learning. The findings of these two studies suggest that teachers' thinking about the purposes of assessment and their assessment practices cannot be predicted just from the assessment purposes and practices mandated or prescribed in the assessment policies of an educational system. The findings also suggest that teachers' thinking about their particular context of assessment and their assessment practices should be explored. This can be done by listening to

the voices of the teachers working in that context and observing what assessment practices they implement in their classrooms.

In Vietnam tests and examinations are very common in education. However, the Ministry of Education and Training's (MOET) initiatives for reform in assessment at the school level have recently laid the foundation for a culture of assessment. The entrenched purposes and practices of testing and the purposes and practices of assessment recently promoted by the MOET may have created a new context of assessment in Vietnam.

1.2. Context of the Study

This section discusses Vietnam's foreign language policies and the development of English language education. The section focuses on the factors that have affected assessment in school education, researchers' calls for reform in assessment, the initiatives for reform in assessment, and the status of assessment in English language education in Vietnam.

1.2.1. Vietnam's foreign language policies.

Foreign language education policies in Vietnam have been affected by the country's political history (Denham, 1992). Chinese, French, and Russian were the dominant foreign languages in Vietnam in different historical periods. The birth of the "Innovation Policy" (*Chính sách Đổi mới*) initiated by the government of Vietnam in 1986 has brought about many changes in diplomatic and economic policies, and these changes have directly influenced the changes in foreign language education policies (Huong, 2010; Minh, 2007; Thinh, 2006). Since the early 1990s, foreign investors have come to Vietnam to do business, thus English proficiency has become essential for individuals to gain employment (Nunan, 2003; Thinh, 2006). As a result of this, the demand for speakers of English has exceeded supply, and the English language learning movement has rapidly developed throughout the country (Thinh, 2006). The status of English as the dominant foreign language has been recognised and supported by the government of Vietnam, and it has been actively promoted through various policies and initiatives.

1.2.2. The growth of English language education in Vietnam.

The rapid development of the English language learning movement throughout the country and the dominant status of English as a foreign language in Vietnam have attracted much attention from educational policy-makers (Nga, 2006). Attention, effort, and investment have been directed to raising the status of English as a foreign language and to improving the quality of English language education through reforms in the curriculum, textbooks, teaching methodology, and

assessment of students' achievement (Canh & Barnard, 2009; Nga, 2006; Nunan, 2003; Thinh, 2006).

However, the goals and expectations of educational policy-makers and textbook writers have not been realised. The inadequacies and/or inappropriateness in training English language teachers, in the provision of professional development programmes (Can, 2007; Canh & Barnard, 2009; Nunan, 2003), in teaching methodology (Canh & Barnard, 2009), and in English language teachers' teaching competence (Canh & Barnard, 2009; Nunan, 2003), especially in assessment (V. T. P. Anh, 2006; Huyen, 2006), have unfavourably affected the quality of English language education in the country. Assessment has been highlighted as an area of particular concern.

1.2.3. Assessment in English language education in Vietnam.

Assessment in school education in Vietnam has been influenced by a number of factors. The first and foremost factor that has directed the practices of assessment in school education in Vietnam is the "achievement disease" (*bệnh thành tích*) (P. S. Anh, 2006; V. T. P. Anh, 2006; Nga, 2006). "Achievement disease" refers to a situation in which test results do not accurately reflect students' learning outcomes because students cheat to obtain high marks, teachers manipulate tests and marks to raise their students' marks, and school administrators either ignore these facts or encourage teachers to raise students' marks. Consequently, false test results are reported to parents and district and provincial educational administrators because teachers and school administrators want to be rewarded and not criticised. To be rewarded and to avoid criticism for poor test results, schools, teachers, and students are said to try their best to do everything they can, including narrowing the curriculum, preparing students for tests and examinations, cheating, and manipulating tests and marks to generate the best marks because marks mean everything, not only to students and their parents but also to teachers and schools (P. S. Anh, 2006; Cuong, 2006; Tuyet, 2006).

The second influence on assessment in schools in Vietnam is the types of tests used for high-stakes tests and examinations. The main form of assessment in English instruction is paper-and-pencil tests which focus on students' memory and recall (V. T. P. Anh, 2006; Huy, 2006; MOET, 2009b). Multiple-choice tests have been welcomed by educational administrators as they believe that multiple-choice tests are more convenient, objective, accurate, able to cover a wider scope of knowledge, are cheaper, and fairer (P. S. Anh, 2006; Cuong, 2006; Tuan, 2006; Tuyet, 2006). Since 2006, tests of English language in the General Education Diploma Examination (GEDE) and the National University Entrance Examination (NUEE) have consisted of 100% of multiple-choice items (MOET, 2005b).

Third, there is a lack of professional development to prepare teachers for change in assessment in Vietnam. Specifically, assessment is not part of teacher training programmes (P. S. Anh, 2006; V. T. P. Anh, 2006; Tuyet, 2006), and in-service teachers lack opportunities for professional development in assessment (Canh & Barnard, 2009). Consequently, teachers lack knowledge and skills in assessment (Cuong, 2006; Nga, 2006; Pham, 2013; Thuan, 2005; Tuan, 2006). This is especially true for English language teachers as the demand for English language teachers has exceeded supply due to the rapid growth of English education in Vietnam in the last 25 years (Canh & Barnard, 2009; Huong, 2010; Nunan, 2003). The presence of these influences suggests that assessment in school education in Vietnam needs reform.

Many authors have appealed for drastic reform in assessment (P. S. Anh, 2006; V. T. P. Anh, 2006; Huyen, 2006; Tuyet, 2006). This is because assessment in school education in Vietnam neither accurately measures students' learning outcomes nor helps improve teaching and learning (P. S. Anh, 2006; V. T. P. Anh, 2006; Huyen, 2006; Nga, 2006). Rather, it has prevented the reform in teaching methodology and has made students passive (V. T. P. Anh, 2006; Huyen, 2006; Nga, 2006). Some authors have suggested specific measures that need to be taken to improve assessment practices in Vietnam. As assessment in Vietnam has been implemented mainly through paper-and-pencil tests for accountability and managerial purposes (P. S. Anh, 2006; V. T. P. Anh, 2006; Nga, 2006; Tuyet, 2006), there is a need to reform the purposes and methods of assessment (Huyen, 2006; Tuyet, 2006). Some authors have suggested that assessment must be for the purposes of improving teaching and learning (Huy, 2006; Huyen, 2006; Mien, 2006; Tuyet, 2006). Several authors have argued that various types of assessment should be used (Huyen, 2006; Tuyet, 2006). Educational policy-makers have positively responded to the calls for reform in assessment in school education in Vietnam.

Over the last 10 years, there have been a number of initiatives to promote reform in assessment in school education in Vietnam. For example, reform in assessment was one of the requirements in the Prime Minister's direction for reform in school education (Prime Minister, 2001). Indeed, since 2002, reform in assessment has been one of the key responsibilities for the provincial Departments of Education and Training (DOETs), school administrators, and teachers across the country (MOET, 2002, 2003a, 2004, 2005a, 2006d, 2007, 2008a, 2009d, 2010a, 2011d, 2012b, 2013, 2014). In the 2007-2008 school year, the MOET first directed the provincial DOETs to direct primary, secondary, and high school teachers to construct tests and item banks and to train teachers in assessment (MOET, 2007). In 2009, the MOET issued a document directing the provincial DOETs to organise conferences on assessment to promote reform in teaching and assessment methods in Literature, History, and Civics in secondary and high schools (MOET,

2009a). Also in 2009, the MOET held a national conference on this topic (MOET, 2009b). Similar conferences on assessment were held in the curriculum areas of Mathematics, Physics, Biology, and foreign languages in the 2010-2011 school year (MOET, 2010a). In addition to these conferences, in 2010, the MOET published a guide for constructing tests and item banks for school administrators and teachers of English language in high schools (MOET, 2010e). Although reform in assessment has been highlighted since the 2002-2003 school year as an important responsibility in schools, not much change has occurred with respect to both the purposes and types of assessment (MOET, 2009b).

Assessment in English language education in Vietnam mainly serves the purpose of summarising student learning outcomes. Although four language skills (i.e., listening, reading, speaking, and writing) and linguistic knowledge of English are expected to be assessed (Minh, 2007; MOET, 2003b, 2010c; Van, Hoa, Loc, Minh, & Tuan, 2006), listening and speaking skills are hardly ever assessed in the classroom (Can, 2007; Canh & Barnard, 2009; Nga, 2006). These two skills are rarely assessed because they are not included in the GEDE, the NUEE, and other high-stakes tests. In effect, grammar, vocabulary, reading, and writing are the foci of testing (Bock, 2000; Canh & Barnard, 2009; Nga, 2006). Moreover, assessment is implemented mainly to get marks for reporting to parents; for ranking, classifying, and certifying students; and for educational administrators to judge the quality of teachers and schools (Nga, 2006). Owing to these purposes of assessment, English language teachers focus on preparing their students to obtain good marks on tests and examinations (Canh & Barnard, 2009; Nga, 2006). However, they are also expected by educational policy-makers to change assessment purposes and to use various assessment methods in their classrooms. This situation may result in teachers' confusion about their contexts of assessment. It may also lead them to negotiate the multiple purposes of assessment in their contexts.

1.3. Statement of the Problem

This section clarifies what a context of assessment is, why it is important to investigate it, and how to explore it. This section also states the problem of the study.

1.3.1. Definition of a context of assessment.

The focus of this current study was to investigate the contexts of assessment in two high schools in Vietnam. Therefore, it is necessary to define what a context of assessment is, to explain why it is worth exploring, and how to explore it.

In defining contexts in education, Blanton and Medina (2009) stated:

A context is constituted of the interweaving of elements mediating human activity, including material, ideal, and social objects; instrumental tools, such as computers, rulers, and pencils;

psychological tools, such as everyday and institutional discourses and cognitive strategies; and rules and regulations, division of labor, participant roles, participation structures, and discourses. (p. 180)

This definition indicates that a context is made up of people, tools, and activities. People use tools to do something. They interact with each other, and they have their own perceptions and intentions, but they have to comply with rules and regulations. In line with this definition, in this thesis, a context of assessment refers to an environment where policy-makers, local educational administrators, school administrators, teachers, students, and parents interact with each other. These people have their own roles, activities, and tools (e.g., policy documents, curricula, textbooks, and tests). They have their own expectations, values, beliefs, and perceptions, but their practices may be restricted by regulations and expectations from other stakeholders. Therefore, to investigate a context of assessment means to examine who are involved in assessment in that context, the roles they take, the tools they use, the actions and activities they carry out, what they intend to achieve when carrying out these actions and activities, and what they value, believe, and perceive related to their assessment practices. When defined in this way, "context" has a very similar meaning to "culture" used in "a culture of testing" or "a testing culture" and "a culture of assessment" or "an assessment culture". These two cultures are identified according to their main purposes, participants, types of assessment, time and frequency of assessment, and the provision (or prohibition) of support and aids in assessment (Inbar-Lourie, 2008a; Kleinsasser, 1995; Lynch, 2001; Wolf et al., 1991) (These two cultures are discussed in the literature review). A context of assessment can be explored by examining policy documents and the stakeholders' perceptions and practices.

1.3.2. The problem of the study.

Principals, school teachers in general, and high school EFL teachers in particular, and students in Vietnam may face challenges in negotiating the traditional purposes and methods of assessment which contrast with those recently initiated and promoted by educational policy-makers in Vietnam.

Principals still play traditional leadership roles, while they are expected to adopt new roles. Particularly, they are partly responsible for overseeing the administration of tests to generate marks for reporting to parents, ranking students, and making decisions about student promotion and retention. At the same time, they are expected to use assessment data for strategic planning and provide teachers with knowledge and skills in using assessment for improving teaching and learning. Research has found that principals perceive that they play various roles in assessment (Renihan & Noonan, 2012) and that reform in assessment creates an additional administrative

workload for them (Ngan, Lee, & Brown, 2010). Principals in contexts where high-stakes testing exists while alternative assessments are promoted for the purpose of improving teaching and learning face challenges in interpreting and using assessment data for these purposes and in supporting their teachers (Ngan et al., 2010). Research has also indicated that the purposes of assessment intended by policy-makers and those perceived by principals are not the same. Principals pay attention to the perceived purposes rather than the intended purposes (Ngan et al., 2010). Principals in high schools in Vietnam are operating in a period of transition and considerable volatility related to assessment reform. They may hold particular perceptions of the contexts of assessment in their schools, and they may have their own ways of negotiating the multiple purposes of assessment in their contexts.

EFL teachers in high schools in Vietnam are required to administer 12 tests per year to generate marks for ranking, retention, and promoting students (MOET, 2006c). They also need to prepare their students for high-stakes tests and examinations (Canh & Barnard, 2009; Nga, 2006). In addition to taking 12 English tests per year as Grade 10 and 11 students, Grade 12 students also take an English examination in the GEDE. Many of the Grade 12 students who have passed the GEDE have to take an English examination in the NUEE. At the same time, EFL teachers are expected to use various methods of assessment, including projects, oral presentations, self-assessment, and peer-assessment, to modify their teaching and to improve students' learning (MOET, 2009d, 2011d, 2014). They are expected to implement these 12 tests, which represent continuous, periodic, end-of-term, and end-of-year tests (MOET, 2011d, 2014).

The students take tests to obtain marks for reporting, ranking, promotion, and retention. At the same time, they are expected to actively participate in assessment in the classroom in order to improve their learning (MOET, 2014). They are expected to demonstrate their learning outcomes, how they learn, and how they apply the knowledge they have learned (MOET, 2014). In addition, they are required to assess themselves and their peers (MOET, 2014).

When principals and EFL teachers are required to implement different types of assessment for multiple purposes, they place emphasis on some purposes of assessment at the expense of others, and they may prioritise some assessment practices and ignore others, depending on their perceptions. Understandings about these participants' perceptions and practices of assessment will help understand how they negotiate the multiple purposes of assessment in their schools. Students will also be required to respond to the various types of assessment and tests they are assigned and their responses to these assessments and tests will be influenced by their perceptions and their views about the expectations of their teachers and parents. Together with the findings about the participants' perceptions and practices of assessment, insights into the ways they negotiate the

multiple purposes of assessment will be identified and these, in turn, will aid in developing an understanding of the contexts of assessment.

Even though parents are involved in their children's learning and assessment and they may influence their children with respect to studying for tests, especially outside of school, they were not directly implicated by policies related to assessment and assessment practices in the schools. Therefore, parents were not included as participants in the present study.

The period of reform in assessment in school education in Vietnam may create a context of assessment which is neither typical of a culture of testing nor typical of a culture of assessment. The problem is that these stakeholders' perceptions of their contexts of assessment, how EFL teachers implement assessment in their classrooms, and how these stakeholders negotiate the multiple purposes of assessment in their contexts are not known. Insights into their perceptions of the contexts of assessment in their schools, classroom assessment practices, and the way they negotiate the multiple purposes of assessment will contribute to knowledge about contexts of assessment in the schools under investigation and in schools which share similar characteristics. These insights could be obtained by examining relevant policy documents, the participants' perceptions of assessment in their schools, and their assessment practices These insights are necessary for policy-makers to design professional development programmes for principals and teachers. They are also helpful for principals, EFL teachers, and students to reflect on their perceptions and assessment practices.

1.4. Aim and Research Questions

The aim of this study was to investigate the contexts of assessment in Grade 10 to 12 EFL classrooms in Vietnam. In order to investigate these contexts, the study addressed the following research questions:

1. What are the perceptions of the contexts of assessment held by the principals, the Grade 10 to 12 EFL teachers, and students in two high schools in Vietnam?
2. How do the Grade 10 to 12 EFL teachers implement assessment in their classrooms?
3. How do the principals, the Grade 10 to 12 EFL teachers, and students negotiate the multiple purposes of assessment in their schools and classrooms?

1.5. Significance

The study is significant for a number of reasons. First, there is limited research that examines principals', high school EFL teachers', and students' perceptions of the contexts of assessment in their schools and how they negotiate the multiple purposes of assessment. A

systematic search for studies of principals' thinking about assessment shows that principals' perceptions of the contexts of assessment in their schools have not received much attention from researchers. A few relevant studies have investigated principals' perceptions of changes in assessment policies (Ngan et al., 2010), how principals use data for decision making (Shen et al., 2010), and principals' perceptions of how data are used and misused in schools (Militello, Bass, Jackson, & Wang, 2013). Several relevant studies have focused on principals' perceptions of their role in large-scale assessment (Newton, Tunison, & Viczko, 2010; Prytula, Noonan, & Hellsten, 2013).

Of the studies about English-as-a-foreign-language (EFL) teachers' perceptions and practices of assessment conducted since 2000, a number of researchers have chosen elementary or primary school teachers as their participants (Butler, 2009; Butler & Lee, 2010; Chan, 2007; Gattullo, 2000). Differences exist in the perceptions and practices of these two groups of teachers. The studies which had EFL teachers in high schools as participants have investigated testing practices in high schools (Pekkanli, 2010) and the factors underlying EFL teachers' perceptions and use of alternative assessments (Inbar-Lourie & Donitsa-Schmidt, 2009).

A number of studies of students' thinking about assessment have been conducted. However, these studies have focused on students' perceptions of high-stakes testing (Triplett & Barksdale, 2005), of assessment in social-studies courses in high schools (Yildirim, 2004), and of assessment in general (Brookhart & Bronowicz, 2003; G. T. L. Brown & Hirschfeld, 2007, 2008; G. T. L. Brown, Irving, Peterson, & Hirschfeld, 2009). Only one study has focused on students' perceptions of assessment in English classrooms, but this study was carried out in a context where English was used as the first language (Moni, van Kraayenoord, & Baker, 2002). There have been very few studies about high school students' perceptions of assessment in English instruction in EFL contexts.

Second, there is evidence that teachers' perceptions and practices of assessment are influenced by the subjects and the grade levels they teach (McMillan, 2001). The current study of Grade 10 to 12 EFL teachers' perceptions of the contexts of assessment in their schools, how they implement assessment in their classrooms, and how they negotiate the multiple purposes of assessment will broaden the body of knowledge about contexts of assessment in a country, Vietnam, undergoing considerable change in assessment in education.

The third reason for carrying out this study is as a response to an appeal from Vietnamese researchers who have called for studies into Vietnamese teachers' perceptions of the purposes of assessment and their assessment practices (Huyen, 2006; Tuyet, 2006). They have argued that insights into teachers' perceptions of assessment purposes and practices are necessary for policy-

makers, teacher educators, and teachers and are beneficial to changing the purposes and practices of assessment in Vietnam (Huyen, 2006; Tuyet, 2006). These insights are essential to support changes of assessment in English language teaching in Vietnam and in similar contexts because "any efforts to change educators' pedagogical practices, whether by mandate or through professional development activities, may be doomed to failure, unless [their perceptions of teaching, the process and purposes of assessment, and the nature of learning] are acknowledged, challenged and eventually changed" (Vandeyar & Killen, 2007, p. 101). Specifically, understandings about principals', EFL teachers', and students' perceptions of the contexts of assessment in their schools; the types of assessment that EFL teachers implement and the way they implement assessment; and the way principals, teachers, and students negotiate the multiple purposes of assessment in their contexts will help policy-makers modify policies and design professional development programmes related to reform in assessment. Additionally, teacher educators take into account EFL teachers' existing perceptions and practices of assessment when they design teacher education programmes for EFL teachers. Furthermore, the findings from the study will be useful for principals and EFL teachers in similar contexts of assessment, in Vietnam and throughout the world, to reflect on their implementation of classroom assessment.

     This chapter presented the background information about the context and the research problem of the study. The aim of the study, the research questions, and the significance of the study were also presented. The next chapter presents the ideas of cultures of assessment and teaching and assessment for learning and of learning. It also reviews the literature on principals', teachers', and students' perceptions of assessment and teachers' assessment practices.

Chapter 2. Literature Review

This chapter comprises the review of the literature. In the first section, the ideas related to a culture of testing and a culture of assessment are discussed. The second section focuses on assessment for learning and assessment of learning. To provide a context for the discussion of assessment for learning, the sociocultural constructivist theory of learning and the characteristics of assessment in a sociocultural constructivist learning environment are discussed.

2.1. A Culture of Testing Versus a Culture of Assessment

Researchers have distinguished two general contexts of assessment, and they call these two general contexts of assessment a culture of testing (or a testing culture) and a culture of assessment (or an assessment culture). These two cultures are differentiated through the major purposes of assessment, the participants directly involved in assessment practices, the methods used to collect assessment data, the timing and frequency of assessment, and the provision (or prohibition) of support and aids that students are allowed to use in assessment (Inbar-Lourie, 2008a; Kleinsasser, 1995; Lynch, 2001; Wolf et al., 1991). These two cultures are reviewed in the following sub-sections.

2.1.1. A culture of testing.

In this culture, tests and examinations are the main tools of assessment. Students' memory and recall are the foci of assessment that is implemented after a period of learning (Beck, 2000; Kleinsasser, 1995; Sadler, 1989). Tests and examinations are conducted mainly to obtain marks to report to parents and other authorities. Marks are also used to rank, classify, certify, or select students and to make students, teachers, and schools accountable (Kleinsasser, 1995; Stiggins, 2002; Wolf et al., 1991). Test results for these purposes need to be accurate and objective (Estaji, 2011; Lee, 2007). Due to these purposes and criteria, external experts or teachers are powerful participants in the culture of testing (Kleinsasser, 1995).

Kleinsasser (1995) has argued that external experts or teachers are in control of and responsible for testing. They construct tests or use commercially prepared tests and keep them uncontaminated (i.e., secret to students). They monitor the test setting to prevent cheating, and they evaluate, score, and report test results. In a culture of testing, students are "powerless victims" (p. 207) who have little or no control over the testing situation and over when and what is assessed.

Students do not even know how the assessment criteria are set and what they mean because the external experts and their teachers do not share the assessment criteria with them.

The purposes of testing in a culture of testing also influence the types of assessment used. Standardised paper-and-pencil tests, typically true or false, multiple choice and completion tests, are preferred because they are appropriate for summarising student learning. Testing requires measuring learning outcomes in an inexpensive, speedy, efficient, accurate, and objective manner (Kleinsasser, 1995; Lee, 2007; Wolf et al., 1991).

Because tests are used to sum up student learning, they are not embedded in everyday instructional activities. Testing is separate from instruction and mostly comes at the end of a unit of study (Beck, 2000; Kleinsasser, 1995). In a culture of testing, a test "signals a closure and a time for conclusions rather than a natural opportunity to pose more questions related to connection making and reflective analysis," and synthesis and revision for a test are "an end-of-unit exercise, not an ongoing learning activity" (Kleinsasser, 1995, p. 207). In some cases, the teacher and students only see the test results and do not have an opportunity to look at the students' errors and successes. Consequently, the teacher has no opportunity "to contest, discuss, or to learn from [his or her] students' performances" (Wolf et al., 1991, p. 47).

The criteria of objectivity and the high-stakes or consequences attached to tests require that students complete tests without any support from peers and teachers. Collaboration between students and the use of aids such as books are therefore considered to be cheating (Kleinsasser, 1995; Wenger, 2009; Wolf et al., 1991).

The characteristics of a culture of testing are quite different from those of assessment in a culture of assessment. The next section discusses the characteristics of a culture of assessment.

2.1.2. A culture of assessment.

Assessment in a culture of assessment also has a number of noticeable features. Assessment in this culture is used to obtain concrete information of what students can and cannot do (Tierney, 2014; Valencia, 2011; Wolf et al., 1991). Such information is used to modify teaching and learning activities to enhance student learning (Black & Wiliam, 1998a; Klenowski, 2009). The main purpose of assessment is to enhance learning rather than to obtain accurate, objective assessment data. Additionally, assessment is a learning event "in which students and teachers might learn, through reflection and debate, about the standards of good work and the rules of evidence" (Wolf et al., 1991, p. 52). Also, through self-assessment and peer-assessment, "students learn to have opinions and make informed, thoughtful judgements" about their own work and that of their peers (Kleinsasser, 1995, p. 208).

Although creating assessment questions and projects and evaluating and judging assessment products are still the teacher's responsibility in this culture, students take a more active role as participants in the assessment process (Dixon & Haigh, 2009; Kleinsasser, 1995). In the culture of assessment, the teacher and his or her students share assessment decisions and assessment criteria (Estaji, 2011; Inbar-Lourie, 2008b; Kleinsasser, 1995; Tierney, 2014). Students may even have some choice about what assessments to complete (Kleinsasser, 1995). In addition to their teacher, students have their peers as an authentic audience for their oral presentations, exhibitions, or performances (Kleinsasser, 1995).

In a culture of assessment, both students' learning processes and their learning outcomes are assessed (Butler & Lee, 2010; Estaji, 2011; Inbar-Lourie, 2008a). These processes rather than their learning products are the focus of assessment (H. D. Brown, 2004; Wolf et al., 1991). Therefore, students are often asked to write, read, and solve problems (Wolf et al., 1991) to "demonstrate what they know and can do in a real-life or meaningful situation" (Kleinsasser, 1995, p. 208). This emphasis means the use of a multitude of assessment types, including not only paper-and-pencil tests and examinations but also oral questions and answers, oral presentations, quizzes, planned and unplanned observations, self-assessment, peer-assessment, practice tasks, assignments, essay tests, term papers, projects, student journals, exhibitions, and portfolios (Klenowski, 2009; McMillan, 2001; Watt, 2005; Wolf et al., 1991).

In terms of time and frequency, assessment is integrated into instruction (Estaji, 2011; Zane, 2009) and implemented through "dialogue, demonstration, and observation" in an ongoing manner (Klenowski, 2009, p. 264). Assessments become episodes or events of learning (Hargreaves, 2005; Kleinsasser, 1995; Wolf et al., 1991) because they provide information that helps modify and enhance teaching and learning (Black & Wiliam, 1998a; Klenowski, 2009).

With regard to the support and aids provided to students in assessment settings, in a culture of assessment, students are encouraged to collaborate with one another and to use other aids such as books and other materials (Baharloo, 2013; James, 2008; Kleinsasser, 1995; Wolf et al., 1991).

The characteristics of a culture of assessment are similar to the characteristics of assessment in the sociocultural constructivist learning environment and of assessment *for* learning, which is discussed in the next section.

2.2. Assessment *for* Learning and Assessment *of* Learning

Assessment *for* learning and assessment *of* learning are differentiated according to the purposes, types, and time of assessment; participants involved in assessment; and the provision or the prohibition of support to students. This section summarises the main features of assessment *for*

learning and assessment *of* learning. The section also argues that assessment *for* learning is typical of the culture of assessment whilst assessment *of* learning is typical of the culture of testing. Prior to discussing assessment for learning, a description is provided of the sociocultural constructivist theory of learning. This theory is presented because the learning environment and practices that emerge from the application of the principles underlying this theory also shape assessment for learning (Serafini, 2001; Stobart, 2008). In particular, the characteristics of assessment in the sociocultural constructivist learning environment are consistent with the central tenets of assessment *for* learning.

2.2.1. The sociocultural constructivist theory of learning.

The sociocultural constructivist theory of learning is grounded in Vygotsky's ideas (Vygotsky, 1978). These ideas have been supported and valued by academics and researchers (James, 2008) because they emphasise the social and cultural factors of learning. This section provides a review of the major ideas of this theory.

While the behaviourist view of learning considers learning as the passive receipt of knowledge from teachers and the cognitive constructivist view sees learning as a function of the individual mind, the sociocultural constructivist perspective views learning as the construction of knowledge and skills through social interactions and cultural mediation (Vygotsky, 1978).

Vygotsky (1978), the originator of this theory, held three main ideas about learning. First, Vygotsky emphasised the importance of social and cultural factors in knowledge construction, saying that learners actively construct their knowledge through interactions with other people in particular contexts rather than through transmission from other people or through the discovery of individuals (Vygotsky, 1978). This notion of learning suggests that assessments may be regarded as learning activities occurring when students work with the teacher and with their peers during assessment.

The second idea in Vygotsky's sociocultural theory is that there is a zone of proximal development (ZPD) (Vygotsky, 1978). The ZDP is defined as "[t]he distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under guidance or in collaboration with more capable peers" (Vygotsky, 1978, p. 86). The notion of the ZDP suggests that collaboration and cooperation are necessary conditions for learning to occur. Vygotsky (1962) asserted this idea when he said, "What the child can do in co-operation today he can do alone tomorrow" (p. 104). The ZDP also suggests that new learning builds on prior knowledge (Baviskar, Hartle, & Whitney, 2009; Pollard, 1990). In the view of sociocultural constructivists, through interactions with others, learners

have "the opportunity to vocalise their knowledge" (Cooperstein & Kocevar-Weidinger, 2004, p. 142), associate old knowledge with new knowledge, and consolidate or get rid of incorrect earlier information and beliefs in order to move forward (Cooperstein & Kocevar-Weidinger, 2004). The idea of the ZDP suggests that support and aids are important for knowledge construction, and that students should be encouraged to use these resources when they are assessed if assessment activities are considered as learning events.

Third, Vygotsky maintained that learning is contextually situated (Vygotsky, 1978), so understanding how learning is happening in an environment is more important than knowing the product of that process (Lutz & Huitt, 2004). This idea suggests that authentic tasks should be valued because meaningful learning develops through authentic tasks (J. G. Brooks & M. G. Brooks, 1999; Lutz & Huitt, 2004). These tasks are defined as activities chosen to simulate those that students may encounter in their real life or assignments (Cooperstein & Kocevar-Weidinger, 2004).

Advocates of the sociocultural constructivist theory of learning have suggested a number of features that would be typical of a classroom where the sociocultural theory of learning is adopted. In such a classroom, the students' role and responsibility are different from those in a traditional classroom. Specifically, they take a central role in their learning (Baviskar et al., 2009; Milbrandt, Felts, Richards, & Abghari, 2004). In addition, they are given complex and challenging phenomena or situations that require them to examine their existing knowledge and structures and to reorganise and construct new models (J. G. Brooks & M. G. Brooks, 1999; Chen, 2003). Furthermore, students do not simply need a good memory of isolated facts to be repeated on tests and assignments; rather, they have to take an active role and responsibility for communicating and justifying their ideas in the classroom (Chen, 2003). They are encouraged to "construct their own views, explore their own interests, engage their own passions, and create newly empowered visions of self and learning in a supportive community" (Milbrandt et al., 2004, p. 33). Owing to these significant changes in students' role and responsibility in the classroom, lessons in sociocultural constructivist classrooms are typically more student-centred than those in traditional classrooms (Milbrandt et al., 2004). Students also have more responsibility and actively participate in assessment (Gipps, 1999; Lynch, 2001; Milbrandt et al., 2004).

In short, most sociocultural theorists have agreed that learning requires students' active participation in social interaction with other people, especially with more capable others (James, 2008; Palmer, 2005), in particular social and cultural contexts in order to construct and reconstruct knowledge (Palmer, 2005). These perspectives on learners and learning have implications for assessment.

2.2.2. Characteristics of assessment in a sociocultural constructivist learning environment.

The sociocultural constructivist theory is not a theory of assessment but it can be applied to assessment because learning theories can have implications for assessment (Fautley & Savage, 2008; James, 2008; Serafini, 2001). This section specifies the typical characteristics of assessment in a sociocultural constructivist environment.

Assessment in a sociocultural constructivist learning environment has a number of identifiable characteristics. First, assessment is an integral part of learning and teaching (James, 2008; Serafini, 2001; Zane, 2009). Assessment is "embedded in the social and cultural life of the classroom" (Gipps, 1999, p. 378), and is carried out by students and teachers to facilitate learning rather than by external assessors and authorities for accountability purposes (James, 2008; Serafini, 2001). Lutz and Huitt (2004) argue that a "principle of the constructivist approach" (p. 87) is that assessment should be carried out as a natural part of the learning process rather than an activity performed at the end of the process. Similarly, Zane (2009) argues that assessments can be deeply embedded into the curriculum so that students do not recognise that they are being assessed. This characteristic is congruent with the sociocultural theory of learning because, in order for assessment to support and improve learning, it must be an event in which students actively interact with their teachers and peers (Nieto, 2010). They use the support of their teachers and peers (James, 2008) and their prior knowledge to change their old ideas or beliefs and to construct new knowledge (Cooperstein & Kocevar-Weidinger, 2004).

The second characteristic of assessment in the sociocultural learning environment is that learners are supported when they are assessed (Elwood & Klenowski, 2002; Gipps, 1999; James, 2008; Lutz & Huitt, 2004). Gipps (1994) suggested that if Vygotsky's idea about the ZDP is valued, students should be allowed to use support and aids when they are assessed because these reduce the students' reliance on memory and increase their thinking and problem-solving ability. Another reason for allowing for support and aids in assessment is that students' performance level on a task varies according to how much support they are given (Suizzo, 2000). According to Gipps (1999), one way to give support to students when they are assessed is to assess them in collaborative groups so that they can give and receive support as well as feedback from one another.

The third characteristic of assessment in the sociocultural constructivist learning environment is that learners take more responsibility and actively participate in the assessment process while teachers play the role of facilitators rather than that of providers and judges of assessment (Gipps, 1999; Lynch, 2001; Milbrandt et al., 2004). This characteristic is consistent with the sociocultural theory and the view that students take a central role in their learning (Baviskar et

al., 2009; Milbrandt et al., 2004). This is because they have to actively participate in interactions to construct knowledge (Kaufman, 2004; Kozulin, 2002; Nieto, 2010). In this assessment environment, students take a key role (Dixon & Haigh, 2009) because they are expected to participate in the process of developing assessment procedures, including sharing the criteria and standards for judging performances (Lynch, 2001). Students make judgments about their learning (Lutz & Huitt, 2004) and are expected to develop skills in self-monitoring and self-regulation to improve their learning (Dixon & Haigh, 2009).

The fourth characteristic of assessment in the sociocultural constructivist environment is that various forms of assessments are used in multiple settings to collect evidence about different aspects of students' attainment (James, 2008; Serafini, 2001; Zane, 2009). Serafini (2001) maintained that both quantitative and qualitative methods should be used to collect information about students' attainment, whilst James (2008) suggested that judgements about students' attainment should be holistic and qualitative rather than atomised and quantified as they are in the measurement approach.

The fifth characteristic is that assessment should be carried out in authentic and meaningful contexts (James, 2008; Serafini, 2001; Zane, 2009). These contexts are often practical, real-world contexts. James (2008) argued that students' abilities and skills should be assessed in situations where complex and situated problems are solved. This is because learning and real-world competence are context-specific and depend on previously experienced contexts. That is, assessment is seen as a social, contextually-specific, interpretive activity (James, 2008; Serafini, 2001; Zane, 2009).

The sixth characteristic is that, for assessment to improve learning, the process of student learning in its social setting rather than the product of this process should be the focus of assessment (Elwood & Klenowski, 2002; James, 2008; Serafini, 2001). Assessment must elicit information about students' learning processes (Serafini, 2001). Particularly, assessment must focus on discovering how well students use available intellectual, human, and material resources to formulate problems, work effectively, and evaluate their efforts (James, 2008). Therefore, assessment in this environment emphasises students' metacognitive skills (Elwood & Klenowski, 2002).

These six characteristics of assessment in the sociocultural constructivist learning environment are also the characteristics of assessment *for* learning, and this is discussed in the next section.

2.2.3. Assessment *for* learning.

A number of authors have provided definitions of assessment *for* learning (e.g., Assessment Reform Group, 2002; Black & Jones, 2006; Black & Wiliam, 1998a; Boud, Cohen, & Sampson, 1999). However, the words in these definitions have often confused educational policy-makers and practitioners (Klenowski, 2009). As a result, there has been a "misunderstanding of the principles, and distortion of the practices, that the original ideals [of these definitions of assessment *for* learning] sought to promote" (Klenowski, 2009, p. 263). In order to avoid misunderstanding of assessment *for* learning, the participants at the Third Conference on Assessment for Learning held in 2009 in Dunedin, New Zealand, offered the "second-generation definition of Assessment for Learning" as follows: *"Assessment for Learning is part of everyday practice by students, teachers and peers that seeks, reflects upon and responds to information from dialogue, demonstration and observation in ways that enhance ongoing learning"* (Klenowski, 2009, p. 264; italics in original). This definition is more specific than those that had been offered before its introduction.

Assessment *for* learning serves the purposes of "enhancing ongoing learning," and thus it is "part of everyday practice" (Klenowski, 2009, p. 264). Assessment *for* learning can be conducted through "dialogue, demonstration and observation" which involve not only the teacher but also the students (p. 264). The features included in this definition are also included in other authors' discussions about assessment *for* learning, and some are summarised below.

Assessment *for* learning is used to gather evidence about student learning in order to modify teaching and to meet student learning needs during instruction (Valencia, 2011). Therefore, assessment *for* learning is integrated in teaching and learning (Bennett & Gitomer, 2009; Stobart, 2008; Weeden et al., 2002) and "implemented on a day-by-day and even minute-by-minute basis" (Valencia, 2011, p. 387). When assessment is an integral part of teaching and learning, multiple methods of assessment can be used.

Assessment *for* learning involves "all those activities undertaken by teachers, and/or by their students" (Black & Wiliam, 1998a, pp. 7-8). Even a formal test administered for summative purposes can also be used for learning if the test is conducted when learning and teaching are occurring (Davison & Leung, 2009). The use of multiple methods of assessment helps assess all aspects of students' achievement (Bennett & Gitomer, 2009; Carr, 2008; Stobart, 2008).

In a context where assessment *for* learning is valued, greater emphasis is placed on students' learning processes rather than their learning outcomes (H. D. Brown, 2004; Stobart, 2008). In such a context, students are actively involved in the assessment process. They share assessment decisions and assessment criteria with the teacher (Kleinsasser, 1995; W. Morgan & Wyatt-Smith, 2000). As a result, they become more responsible for their learning (Beck, 2000; Black, 1998;

Stobart, 2008) and learn about standards of good work (Stiggins, 2005; Wolf et al., 1991). Moreover, students are allowed to collaborate with their peers and teacher because it is believed that when students and teachers work together, they know what has been learned, what needs to be improved, and how to improve it (Beck, 2000).

These characteristics of assessment *for* learning are also found in discussions about *formative* assessment because assessment *for* learning and formative assessment are often used as synonyms (Crooks, 2011; Lee, 2007). However, some authors have argued that they should be differentiated (Black, Harrison, Lee, Marshall, & Wiliam, 2003; Davison & Leung, 2009; Stiggins, 2005). There are two reasons for the differentiation of formative assessment and assessment *for* learning. First, some authors have argued that they have different meanings. While formative assessment informs the teacher about student achievement, assessment *for* learning informs both the teacher and the students about their learning (Stiggins, 2005). In addition, formative assessment means providing more frequent assessment while "assessment FOR learning focuses on day-to-day progress in learning . . ." (Stiggins, 2005, p. 328; emphasis in original).

Second, there have been misinterpretations of the term "formative assessment". For example, formative assessment is mistakenly equated with continuous or alternative assessment, even though such alternative assessments as oral presentations and portfolios can also be used for summative purposes (Davison & Leung, 2009). Additionally, formative assessment is mistakenly equated with informal assessment, although formal tests can be used to improve learning and informal assessments can be used for summative purposes (Black et al., 2003). In order to avoid these misinterpretations, assessment *for* learning should be used instead to refer to any assessment, including both traditional formal tests and informal alternative assessments, which is used to collect evidence of student learning in order to modify teaching and enhance student learning (Black et al., 2003; Davison & Leung, 2009).

The characteristics of assessment *for* learning are quite different from those of assessment *of* learning, which is discussed in the next section.

2.2.4. Assessment *of* learning.

Assessment *of* learning, also referred to as summative assessment (Crooks, 2011; Davison & Leung, 2009; Lee, 2007), is typical of assessment in a culture of testing. It is used to sum up students' achievement over a period of time (Beck, 2000; Sadler, 1989) by objectively and accurately measuring their knowledge and skills according to standards in standardised conditions (Lee, 2007). Assessment *of* learning is used for a variety of purposes and mainly for administrative and reporting purposes (Delandshere, 2002; Genesee & Upshur, 1996). For example, it is used to

compare students with one another, to review their progress, to provide information about their abilities when they transfer to other classes or schools, to certify them for further or higher education or employment, and to judge the effectiveness of teachers and schools for accountability (Black, 1998; Stiggins, 2002; Valencia, 2011). To be aligned with these purposes, tests need to be objective, accurate, efficient, and economical to administer and to score (Lee, 2007; Valencia, 2011). They often cover a wide range of content (Valencia, 2011). As a result of these criteria, high-stakes, standardised multiple-choice tests are often conducted at the end of a learning unit (Beck, 2000). In situations of assessment *of* learning, students are passive test-takers or not allowed to collaborate with peers and to use support and aids in testing settings (Kleinsasser, 1995). Although summative tests results are not accurate evidence of student learning (R. L. Linn, 1986; Valencia, 2011), they play an important role in policy, curricular, and classroom decisions and result in a multitude of undesirable effects (R. Linn, 2000; Valencia, 2011).

The negative consequences of summative tests, especially those associated with high-stakes testing, have been indicated by many authors (e.g., Hoffman, Assaf, & Paris, 2001; Shepard, 2000). Among the negative effects of summative assessment are the following: spending too much time preparing students for tests and examinations, especially for those who are close to passing, while excluding low-achieving students from testing; narrowing the curriculum; and focusing on teaching and learning basic skills (Allington & McGill-Franzen, 1992; Koretz, 2008; R. Linn, 2000; Nolen, Haladyna, & Haas, 1992; Shepard, 2008). In addition, summative assessment creates anxiety and worry and discourages students, rather than supports, directs and encourages them (Black & Wiliam, 1998a; Shepard, 2000; Stiggins, 1999).

Although assessment *of* learning does more harm than good, it has been pervasively applied (Stiggins, 1999). The reasons for its pervasiveness are many and varied. One reason is that students' scores in large-scale standardised tests are used as a measure to make teachers and students accountable (Stiggins, 1999). Other reasons may include teachers' lack of knowledge and skill in assessment (Pryor & Lubisi, 2002) and teachers' lack of time and resources for formative assessment (Kanjee, 2009; Watt, 2005). Assessment *of* learning is also said to be necessary in contexts where tests and examinations are used to "ration access to scarce opportunities" (Pryor & Lubisi, 2002, p. 679).

Up to this point in the chapter, the researcher has discussed the characteristics of a culture of testing and a culture of assessment. Moreover, assessment *for* learning and assessment *of* learning have been reviewed. The review argued that the characteristics of assessment *for* learning are typical of a culture of assessment. In contrast, the characteristics of assessment *of* learning are typical of a culture of testing.

It can be said that a culture of testing, which focuses on assessment of learning, and a culture of assessment, which focuses on assessment for learning, reflect in general different contexts of assessment. These general contexts of assessment are distinguished according to their components (i.e., participants, values, practices, and tools). In a culture of testing, the participants are more likely to be external experts, administrators, and teachers; the values are more likely to emphasise objectivity and accuracy, and focus on outcomes related to individual ability and the products of learning. The tools used in a culture of testing are likely to be paper-and-pencil tests and examinations. In contrast, in a culture of assessment, the participants are likely to be the teachers and their students; and the values are more likely to emphasise the provision of feedback to enhance learning. Student self-regulation of learning, awareness, and monitoring of learning are also important in a culture of assessment. Practices that develop these latter skills and abilities are fostered in a culture of assessment.

2.3. Language Instruction and Assessment

Language teaching and learning are different from teaching and learning in other subjects such as Geography and Mathematics. While language is a means or medium for the teaching and learning of these subjects; in language instruction, language is both the medium and an end of teaching and learning. Therefore, the goals of language instruction and the methods used to observe whether and to what extent these goals have been achieved are different from those of instruction in other subjects. Hence, a review of the literature on language instruction and assessment in language instruction is part in this thesis.

It should be noted that theories about second language (L2) instruction and foreign language (FL) instruction can both be used to discuss EFL instruction and assessment. The justification for this argument is that while there are "important differences" between English-as-a-foreign language (EFL) learners and English-as-a-second language (ESL) learners, there are "obvious parallels" between them because "all newcomers to a language system have to learn its systems of pronunciation, grammar, vocabulary, and socially acceptable ways of using the language" (Kopriva, 2008, p. 187). In addition, methods and approaches to L2 instruction are also often used in FL instruction. As a result, the review in this section uses the literature on both L2 and FL instruction and assessment.

2.3.1. L2/FL instruction.

Dozens of methods of and approaches to instruction have been introduced and applied to L2/FL instruction (H. D. Brown, 2007; Fotos, 2005; Richards & Rogers, 2001; Xia, 2014).

However, the Grammar-Translation Method (GTM) and the Communicative Language Teaching Approach (CLT) have been used as two main L2 and FL teaching methods around the world (Hinkel, 2005). The shift from GTM to CLT was influenced by the inability of GTM to prepare language learners to communicate successfully in real-world contexts (Savignon, 2002). Indeed, the shift from GTM to CLT is the shift from L2/FL instruction that focuses on linguistic forms to L2/FL instruction that focuses on linguistic functions (Alemi, Eslami, & Rezanejad, 2014; Kavanagh, 2012; Kopriva, 2008). In GTM, language learners are expected to master linguistic forms, including morphemes, words, sentences, and the rules for combining them (H. D. Brown, 2007). In CLT, learners are expected to develop the ability to use language for meaningful interactive purposes within various social contexts (H. D. Brown, 2007).

Because linguistic competence, defined as "knowledge 'about' language forms" (H. D. Brown, 2007, p. 219), is insufficient for effective communication in real-world contexts (Bagarié, 2007; H. D. Brown, 2007), language learners are required to develop communicative competence or the ability to use language in social contexts and to conform to sociolinguistic norms appropriately (Savignon, 2002).

Several models of communicative competence have been introduced by Canale and Swain (1980, 1981), Canale (1983), and Bachman and Palmer (1996) (See Bagarié (2007) for a review of these models). The model proposed by Canale and Swain (1980, 1981), which was then modified by Canale (1983), is said to be the most influential model of communicative competence in L2/FL instruction and assessment (Bagarié, 2007). This model consists of four areas of competence: grammatical/linguistic, sociolinguistic, discourse, and strategic competence.

Grammatical competence refers to "knowledge of lexical items and of rules of morphology, syntax, semantics, and phonology" (Canale & Swain, 1981, p. 32). Sociolinguistic competence is defined as knowledge about rules that "specify the ways in which utterances are produced and understood *appropriately*" in communicative events (Canale & Swain, 1981, p. 33, emphasis in original). This competence requires that L2/FL learners know the social rules when they use language in different social contexts (Savignon, 2002). Sociolinguistic competence is also called pragmatic competence, which is concerned with "who can say what to whom, when, and in what manner" (Kopriva, 2008, p. 184). Discourse competence is related to "cohesion (i.e., grammatical links) and coherence (i.e., appropriate combination of communicative functions) of groups of utterances" (Canale & Swain, 1981, p. 33). This is the ability to combine sentences or utterances to form a text as a meaningful whole (Savignon, 2002). Strategic competence comprises both verbal and nonverbal communicative strategies that language learners can use to compensate breakdowns in communication due to their insufficient competence (Canale & Swain, 1981). Strategic

competence helps language learners to enhance the effectiveness of communication (Canale, 1983) by using such strategies as repetition, reluctance, and guessing (Bagarié, 2007) to make corrections or changes or to clarify what has been said (Kopriva, 2008).

In order to develop communicative competence, language learners need to be owners of learning activities, to be active participants in these activities, and to interact with teachers and peers using the target language in more authentic contexts (Agbatogun, 2014). Therefore, CLT, which adopts communicative competence as its "central theoretical concept" (Savignon, 2002, p. 1), emphasises language learners' engagement in activities that require them to use the target language in meaningful and authentic contexts (Kavanagh, 2012). In CLT, learners are often asked to work in pairs or small groups using the target language to play games, role-play, complete tasks, or solve problems (Agbatogun, 2014).

In short, the emphasis on language learners' mastery of linguistic competence in GTM is not advocated in L2 and FL instruction because learners with a good grammatical or linguistic competence may be unable to use the target language to achieve intended functions. Because the ability to communicate in real-world situations is the ultimate goal of language instruction (H. D. Brown, 2007), in CLT, learners are expected to develop communicative competence, which includes not only grammatical competence but also sociolinguistic competence, discourse competence, and strategic competence.

2.3.2. Assessment in L2/FL instruction.

The change in how language proficiency is viewed and how language has been taught have also resulted in a shift in the focus of assessment and how language assessment is administered (H. D. Brown, 2004; T. McNamara & Roever, 2007).

In GTM, language proficiency in a second or foreign language is seen as the mastery of the formal systems of the language, thus assessment focuses on students' grammatical or linguistic competence (H. D. Brown, 2004; Chalhoub-Deville & Deville, 2005; T. McNamara & Roever, 2007). In CLT, communicative competence (H. D. Brown, 2004) and face-to-face communication in real-life situations are valued (Chalhoub-Deville & Deville, 2005; T. McNamara & Roever, 2007). Therefore, students' communicative competence is the focus of assessment in CLT (H. D. Brown, 2004; Chalhoub-Deville & Deville, 2005; T. McNamara & Roever, 2007).

In GTM, language skills (listening, speaking, reading, and writing) and various units of language or discrete points of phonology, graphology, morphology, lexicon, syntax and discourse are assessed separately (H. D. Brown, 2004). Therefore, knowledge-oriented tests (i.e., tests that measure students' knowledge about language) are preferred (T. McNamara & Roever, 2007;

Savignon, 2002). However, decontextualised paper-and-pencil tests do not elicit students' communicative competence because they do not involve them in requesting, responding, and interacting by combining listening and speaking or reading and writing (H. D. Brown, 2001).

In CLT, not only is grammatical competence assessed but features of discourse, sociolinguistic rules of appropriateness, and communicative strategies need to be assessed in order to judge students' communicative competence and help them develop it (Savignon, 2002). To tap into students' communicative competence, teachers should assess their students in authentic contexts where forms and functions are integrated as naturally as possible (Kopriva, 2008). Therefore, face-to-face interaction, especially role-play, is preferred in the assessment of students' language use because students' communicative competence can only be assessed in contexts similar to real-life contexts (T. McNamara & Roever, 2007). Therefore, other performance-based assessments, such as oral presentations, essay writing, open-ended responses, group performance, and other interactive tasks are preferred (H. D. Brown, 2004).

Performance-based and integrative assessments need to be administered to assess students' communicative competence, the desirable goal of language learning, rather than tests of their knowledge of discrete linguistic forms (Kopriva, 2008). However, research has indicated that in contexts where CLT is promoted, discrete-point testing is pervasive for various reasons. For example, in China, high school EFL teachers mainly use multiple-choice tests to assess students' vocabulary and grammar because high-stakes tests include mainly grammar and vocabulary, even though policy-makers expect EFL teachers to develop students' ability in language use (i.e., communicative competence) (Watanabe, 2004). In Japan, high school EFL teachers do not assess students' speaking because it is not included in the university entrance examination. Rather, their assessment focuses on vocabulary and grammar because these are included in the national examination (Kavanagh, 2012; T. McNamara & Roever, 2007), even though students' communicative competence in English is prioritised by policy-makers (Kavanagh, 2012).

In Vietnam, policy-makers expect EFL students to develop communicative competence, and CLT has been adopted. Moreover, English tests are high-stakes for high school students in Vietnam because marks generated from formal tests and examinations are used for decisions related to students' promotion, retention, and university admission. In such a context, EFL teachers in Vietnam may prioritise assessing communicative competence or preparing their students for high-stakes tests and examinations. In addition, their perceptions of assessment may be different from their assessment practices, and they have to negotiate different purposes of assessment in their schools. These issues deserve investigation because understandings about these issues help understand their contexts of assessment.

This section addresses language instruction and assessment. The next sections review research studies about principals', teachers', and students' perceptions of assessment and teachers' assessment practices.

2.4. Perceptions and Practices of Assessment

The review of research studies that have investigated perceptions and practices of assessment indicated that the majority of such studies have focused on a single group of participants. That is, the studies have been conducted to examine the perceptions and/or practices of assessment of either principals (or school administrators), teachers, or students. Studies that focused on more than one group of participants such as school administrators and teachers (Militello et al., 2013) and teachers and students (Segers & Tillema, 2011; Yildirim, 2004) are rare. Therefore, the review of the literature of research studies of perceptions and/or practices of assessment is organized according to the three groups of participants that were of interest in this study (i.e., principals or school administrators, teachers, and students). Studies that have examined parents' perceptions of assessment are not reviewed because parents were not included as participants in the present study.

2.4.1. Principals' perceptions of assessment.

A few studies have investigated principals' perceptions of assessment. Among these studies, some have investigated principals' perceptions of large-scale assessment reform (Prytula et al., 2013) and principals' perceptions of the use of large-scale assessments and their roles related to these assessments (Newton et al., 2010). Research has also investigated principals' perceptions of their roles as leaders in assessment (Renihan & Noonan, 2012). Some studies have examined principals' perceptions of how assessment data were used in their schools (Militello et al., 2013; Shen et al., 2010). In addition, a study has been carried out to investigate principals' perceptions of changes in policies on assessment and school evaluation (Ngan et al., 2010). These studies are reviewed in this section.

2.4.1.1. Principals' perceptions of large-scale assessment.

In order to examine the perceptions that the principals held of large-scale assessment reform and how assessment reform affected their roles as principals, Prytula, Noonan, and Hellsten (2013) used survey questionnaires comprising both closed and open-ended questions. Ninety principals in two rural school jurisdictions and three urban school jurisdictions in Saskatchewan, Canada responded to the questionnaires. Saskatchewan schools participated in the Pan-Canadian Assessment Program (PCAP), the Program for International Student Assessment (PISA), and the

Progress in International Reading Literacy Study (PIRIL). In addition to these national and international tests, schools participated in provincial tests which were aligned with the provincial curriculum and which were intended for gathering information about student learning for improving student achievement in maths, reading, and writing.

The principals in the study perceived that large-scale assessments created pressure to improve teaching and learning and focused on setting student learning goals at the school level. Some principals believed that the large-scale assessments helped school administrators and staff to identify and focus on students' weaknesses. However, a number of principals thought that these assessments were inconvenient and irrelevant and could not be used to improve student learning. Rather, the principals believed these assessments created additional administrative tasks and additional pressures from higher authorities. The principals who perceived that the large-scale assessments had no effect on their roles were not interested in the assessments themselves nor the data gathered from them. These principals reported that they did not use the data for changes in their schools.

The principals in Prytula, Noonan, and Hellsten's (2013) study place more weight on the provincial assessments. Many of them perceived that the provincial assessments positively influenced teaching and learning in their schools. They reported that these assessments influenced decision making, priority setting, and planning. They believed that the provincial assessments helped staff set goals to drive learning improvement. However, a few principals perceived that the provincial assessments were inconvenient and inappropriate for students and staff. Although some principals reported that they used the data to set goals, they believed the assessments themselves had no influence on teaching and learning in their schools. Several principals perceived that the assessments simply confirmed something that staff already knew. Other principals believed that the assessments caused teachers to teach to the test.

The findings of this study indicated that the principals held both positive and negative perceptions of different types of large-scale assessment. However, they valued assessments that were aligned with the curriculum that was used in their schools.

Newton, et al. (2010) investigated elementary principals' perceptions and the use of large-scale assessments in their schools as well as the principals' roles related to these assessments. Specifically, the study focused on principals' perceptions and use of the Canadian Achievement Test (Version 3) (CAT3), the Canadian Test of Cognitive Skill (CTCS), and the provincial achievement tests (PAT). These tests were used in all the schools in a large urban school division in Western Canada. Semi-structured interviews were conducted individually with 25 elementary school principals to generate data for the study.

The principals perceived that large-scale assessment data helped them identify their priorities for school-based planning and priority setting. They reported identifying long-term trends within the student population by analysing the assessment results across a number of years. In addition, they perceived that the large-scale assessment data helped them support their teachers by targeting professional development.

The principals believed that each large-scale assessment was intended for a different purpose. They believed that the data obtained from the CAT3 were useful for school strategic planning. Some principals stated that the CAT3 helped them identify students' strengths and weaknesses. However, some principals thought that the data obtained from the CAT3 were not useful because they provided the results in only a few subject areas.

The principals believed that the data obtained from the CTCS also confirmed what they already knew about individual students' performance. Some believed that the CTCS data helped teachers see discrepancies between students' achievement in classroom assessments and their performance on the CTCS. Some principals believed that the CTCS data were used only for placement into special programmes or high school placement.

Most principals perceived that the PAT was the most significant large-scale assessment for school-level planning because it was based on the province's curriculum and designed by the teachers in the province. Several principals believed that the PAT data provided the most useful evidence of student learning, and they spent more time looking at the data obtained from the PAT for school-level planning.

The principals perceived that data from large-scale assessments were primarily used for accountability purposes. They believed that the data were used to evaluate the effectiveness of the administrative and instructional staff. They did believe, however, that accountability was necessary. But, several principals argued that the division should focus more on formative assessment rather than large-scale summative assessments. They argued that publishing large-scale assessment results would be detrimental to low-achieving schools.

The principals perceived that large-scale assessments fundamentally changed their role. They stated that their primary responsibility was to manage the data, interpret results, incorporate results into strategic planning, and monitor progress. Some principals suggested that they were responsible for ensuring that their teachers complied with and engaged in plans developed from assessment data. In addition, many principals perceived that one of their fundamental roles was to assist teachers in interpreting and contextualising large-scale assessment results. The principals expected their teachers to use assessment data to improve their teaching and student learning in the

classroom. To that end, they had to collaborate with and support teachers in interpreting the data and planning their instruction using the data.

About one-third of the principals admitted that they had limited knowledge of how to interpret results from the large-scale assessments. Several principals perceived that their teachers were not well prepared in interpreting and using assessment data. They also thought that they themselves and their teachers needed support in using assessment data for strategic planning at the school level.

Several principals expected the assessment results to be accurate measures of student achievement. They believed that accuracy would be improved when students' test-taking skills improved. Therefore, these principals believed that teachers needed additional support to prepare students to compete these large-scale assessments.

The findings of this study indicated that the principals held different, even contradictory, perceptions of large-scale assessments in their contexts. The principals felt that their role was one of managing and interpreting data. At the same time, they believed that they had to support their teachers with interpreting the data for the teachers' instruction. Furthermore, the principals admitted that they themselves and their teachers had difficulty in interpreting and using large-scale assessment data for strategic planning. The principals valued assessment data generated from tests that were aligned with the curriculum and from formative assessment because they perceived that data gathered in this way were more relevant to students.

The findings reviewed in this sub-section showed that the principals in these studies believed that different large-scale assessments served different purposes. They valued large-scale assessments that were aligned with the curriculum used in their schools. In addition, different principals held different perceptions of the purposes of large-scale assessments. In general, however, they perceived that large-scale assessment data helped them set priorities and plans for their schools. They also believed that the results of large-scale assessments were used for accountability purposes. Principals thought that large-scale assessment changed their roles. Their role involved interpreting assessment data for strategic planning. However, many perceived that they lacked knowledge and skills for assisting their teachers in using the data to plan for the teachers' instruction. The studies about principals' perceptions of their roles in assessment are reviewed in the following sub-section.

2.4.1.2. Principals' perceptions of their leadership roles in assessment.

In order to investigate principals' perceptions of their assessment leadership role in rural schools, Renihan and Noonan (2012) conducted three focus group discussions with 12 principals in

12 small village schools in three school divisions in Saskatchewan in Canada. In Saskatchewan, there has been an increasing emphasis on large-scale assessment, and data-driven school decision-making although assessment for learning has also been emphasised strongly.

The majority of the principals reported that they played multiple roles of teachers, managers, and supervisors. Some saw themselves as teachers first. Some believed that they should manage assessment professionally and collegially rather than bureaucratically. The principals perceived that they should maintain a balance between ensuring accountability and quality control and nurturing professional empowerment among teachers. Most principals perceived that their roles in assessment included providing teachers with a clear vision and direction for instruction and assessment and promoting discussion of the summative and formative purposes of assessment among staff. The principals reported providing professional development based on the instructional and assessment needs identified by their teachers.

The principals acknowledged that rural schools had some unique features such as parent involvement, an influence from parent values, and community expectations. They also suggested that rural school principals spent more time and energy on administrative issues (e.g., school organisation) and less time on leadership functions such as assessment leadership. The principals in some schools questioned the value of their involvement in large-scale assessment. These principals believed that classroom teachers had the ability and responsibility to implement and use current assessment and grading practices.

The principals perceived that support from central administrators influenced their assessment leadership and knowledge and skills in assessment. They indicated that they needed clear guidance and expectations, professional development, and resources.

Policy-makers in many contexts have reformed assessment and evaluation policies, and principals as assessment leaders in these contexts may have faced challenges created by these policy changes. In order to examine how principals dealt with assessment and evaluation policy changes, Ngan, Lee, and Brown (2010) conducted 23 semi-structured interviews with 14 primary and nine secondary principals from 18 school districts in Hong Kong.

According to Ngan, Lee, and Brown (2010), in Hong Kong, at the time of their study, responsibility for school quality had shifted from central authorities to schools. This shift occurred by promoting internal quality assurance through school-based management, parents' and teachers' participation, and school self-evaluation. They introduced the School Self Evaluation (SSE) and the External School Review (ESR), which were created to assess school quality. This shift made school leaders responsible for school-based improvements. In addition, to mitigate the negative effects of the over-reliance on public examinations, the Hong Kong Education Bureau (EDB) introduced

assessment for learning. The purposes of assessment for learning comprised assessing higher-order thinking skills, providing students with constructive feedback, and sharing learning objectives for the purpose of improving classroom teaching and learning. The new assessment reform made school leaders responsible for much more complex school-based assessment practices compared with their management of the public end-of-year examinations to determine student learning outcomes.

Most principals in this study had negative views about the policy for school quality assurance. They perceived that the EDB used the SSE and the ESR for school accountability. They thought that the SSE and the ESR were not fair or accurate assessments of school quality because the set of criteria used to evaluate schools was narrow, limited, and was not sensitive to the unique features and characteristics of each school in relation to its teachers, students, and school background. They perceived that the quality assurance mechanism was results-oriented and ignored the student intake, and school background. Most principals perceived that the reform was full of contradictions. They thought that the EDB promoted school-based management but it strictly controlled school administration. Many principals perceived that the new assessment policy created an excessive workload and placed unnecessary stress on teachers and students. The principals perceived that the SSA and the ESR were used to close schools that did not achieve well, rather than helping them improve.

The principals perceived that preparing students for the tests was necessary, and they reported familiarising students with the format of the tests and the tested skill (i.e., writing). A number of the principals reported that their schools used drill exercises to help students obtain higher scores. Such activities were more closely related to the principals' personal job security and the survival of their schools than furthering the students' learning.

The review of the studies in this sub-section showed that principals perceived that they played multiple roles in assessment in their schools. In addition, their roles as assessment leaders were influenced by contextual factors. The findings of the studies reviewed in this sub-section also indicated that the purposes of assessment that were intended by policy-makers and those that were perceived by principals were different. When this happened, principals responded to the purposes that they perceived rather than the purposes intended by policy-makers. This finding suggested that research should investigate how principals negotiate the multiple purposes of assessment in their contexts. This will be explored in the present study.

2.4.1.3. Principals' perceptions of assessment data use.

Some researchers have examined principals' perceptions of how assessment data were used and misused. For example, Militello et al. (2013) examined the perceptions of 28 elementary principals in the USA. They were asked to rank order 23 statements to indicate how they used assessment data. After they had sorted the statements, they were also asked to complete a questionnaire. The questions in the questionnaire elicited the participants' reasons for ranking the statements in the order that they chose.

The principals reported collecting, analysing, and using assessment data to make plans for teaching and learning improvement and to evaluate school programmes. They also reported showing their teachers how to use assessment data to inform their instruction. The principals used data to judge students' past performance and predict their future performance. They also used data to identify at-risk students. They reported that they needed knowledge and skills in using data so that they could support their teachers to use data to inform instruction.

A similar study was conducted by Shen and colleagues (2010). They investigated the data that principals used to make decisions and how data informed decision-making in their schools. Shen, at al. (2010) argued that there were three streams of data that could be used for decision making at the school level. These included data about students and community (i.e., demographic descriptive background information about students, teachers, and schools), data about school processes (i.e., what school administrators, teachers, and students do to get results), and student achievement data (i.e., assessment results). In this study, the researchers individually interviewed 16 principals in four urban school districts in Michigan, the USA. In each school district, two elementary principals, one middle school principal, and one high school principal were interviewed.

Of the three streams of data, student achievement data were used most frequently for decision-making, whereas student and community background data and school process data were rarely used. Among student achievement data, results from standardised tests were most frequently used.

The principals used student achievement data for various purposes. First, almost all the principals reported using student achievement data for making teachers accountable, and this was regarded to be the predominant purpose. Second, student achievement data were used to collaborate with parents in helping students learn better. Third, some principals reported using this stream of data to identify students' knowledge gaps. Fourth, student achievement data were used to compare student achievement with specific norms and to identify students' achievement growth. Fifth, student achievement data were used to make decisions directly related to curriculum and instruction

(i.e., grouping and placement, identifying students' weaknesses according to state and other standards, assessing students' proficiency against curriculum objectives, and modifying instruction).

The findings from the studies reviewed in this sub-section showed that principals perceived that assessment could be used to inform instruction and make plans for improving teaching and learning. In addition, they believed that assessment data were used to make teachers accountable for their teaching. The findings in these studies also revealed that student achievement data gathered through standardised tests were the main data source that principals used to make decisions in their schools.

The review in this section indicated that principals in different contexts held various perceptions of the purposes of assessment. The purposes of assessment perceived by principals may be inconsistent with the purposes intended by policy-makers, and principals dealt with assessment in their contexts according to their perceptions rather than according to policy-makers' expectations. In addition, principals perceived that they played different roles in assessment in their schools. Principals in some contexts perceived that they lacked knowledge and skills in interpreting and using assessment data for improving teaching and learning in their schools.

In high schools in Vietnam, where testing is high-stakes but assessment for learning has been promoted, principals may play various roles and they may face challenges in negotiating multiple purposes of assessment. Therefore, a study of principals' perceptions of the contexts of assessment in their schools and how they negotiate the multiple purposes of assessment will extend knowledge about principals' perceptions of assessment. Such a study may be useful for policy-makers in providing principals with knowledge and skills for improving assessment in their schools.

2.4.2. Teachers' perceptions and practices of assessment.

The review of research studies about teachers' perceptions of assessment and their assessment practices is organised in two sub-sections. The first sub-section reviews studies about teachers' perceptions of assessment. The second sub-section reviews studies about teachers' assessment practices.

2.4.2.1. Teachers' perceptions of assessment.

Since 2000, studies about teachers' perceptions of assessment have focused on three major topics. The majority of studies have investigated teachers' perceptions of the purposes of assessment. Some studies have investigated teachers' perceptions of assessment of learning, and others investigated teachers' perceptions of classroom assessment.

2.4.2.1.1. Teachers' perceptions of the purposes of assessment.

Many studies about teachers' perceptions of assessment have examined teachers' perceptions of the general purposes of assessment. Many have used T. G. L. Brown's (2004; 2006) model of the general purposes of assessment. G. T. L. Brown (2004) designed a 50-item questionnaire, called the COA-III (Conceptions of Assessment), to examine the perceptions of assessment. The questionnaire was completed by 525 teachers and administrators from over 290 primary schools in New Zealand. The teachers were asked to indicate the extent of their dis/agreement with the statements related to four major purposes of assessment. These purposes were *Student Accountability* (i.e., assessment made students accountable for their learning), *School Accountability* (i.e., assessment made teachers and schools accountable), *Improvement* (i.e., assessment helped teachers and students improve teaching and learning), and *Irrelevance* (i.e., assessment was used but ignored; assessment was bad for students and teachers). The teachers in the study believed that assessment improved learning and teaching. They also believed that assessment made teachers and schools accountable, but they did not believe that assessment made students accountable or was irrelevant to teachers' work and students' life.

G. T. L. Brown's (2004) questionnaire was modified, resulting in a 27-item questionnaire, called the CoA-IIIA (G. T. L. Brown, 2006). The CoA-IIIA was used to examine the perceptions of the purposes of assessment held by teachers in such contexts as New Zealand (G. T. L. Brown & Harris, 2009), New Zealand and Queensland (G. T. L. Brown & Lake, 2006), Hong Kong (G. T. L. Brown, Kennedy, et al., 2009), and the Netherlands (Segers & Tillema, 2011). These studies found that teachers in these contexts held multiple perceptions of the purposes of assessment. However, the extent to which they agreed or disagreed with each of the four purposes was different. For each example, teachers in New Zealand, Queensland, and Hong Kong agreed most with the *Improvement* purpose and rejected the *Irrelevance* purpose. There was greater agreement with the *Student Accountability* purpose over the *School Accountability* purpose in all three groups. However, the Hong Kong teachers agreed more with the *Student Accountability* purpose and the association between the *Student Accountability* purpose and the *Improvement* purpose was also stronger. In other words, the Hong Kong teachers believed that assessment improved learning because it made students accountable for their learning. Even teachers in the same country but teaching at different school levels responded differently with regard to the extent of agreement with these purposes of assessment. For example, primary teachers and secondary teachers in Queensland expressed different views on the extent of their agreement with the *Improvement* purpose and the *Student Accountability* purpose (G. T. L. Brown, Lake, & Matters, 2011). Specifically, the Queensland primary teachers in the study moderately agreed with the *Improvement* purpose whilst the

secondary teachers agreed more with the *Student Accountability* purpose (G. T. L. Brown et al., 2011). The different degrees of agreement with these purposes of assessment among different groups of teachers were attributed to the different contexts of assessment. In Hong Kong, where assessment often had high-stakes and students learned in a very competitive context, assessment was believed to make students accountable, whilst assessment in New Zealand did not have high-stakes, and the New Zealand teachers did not put much weight on the *Student Accountability* purpose (G. T. L. Brown, Kennedy, et al., 2009; G. T. L. Brown & Lake, 2006). These studies found that teachers fulfilled multiple purposes and their perceptions of these purposes were context-dependent. However, these studies did not investigate how the teachers in these contexts implemented assessment in their classrooms and how they negotiated the multiple purposes of assessment in their contexts. Apart from the studies that used G. T. L. Brown's (2004, 2006) model of assessment purposes, some other studies have also examined teachers' perceptions of the purposes of assessment.

In a context where assessment of learning is dominant, teachers tend to value the summative purposes of assessment. This was the finding in a study in Hong Kong by Cheng (2006). Cheng interviewed eight teachers in four secondary schools in order to explore their perceptions of assessment. They reported that they administered assessment to fulfil parents' expectations and to call for their support. In addition, they believed that assessment informed administrators about teachers' teaching and students' learning. They thought that assessment helped them judge their students' improvement and assessment improved teaching and learning. They reported using alternative assessments such as quizzes, assignments, concept mapping, projects, self-assessment, and peer-assessment. However, they associated assessment with tests and examinations, thus believing that alternative assessments should not be the major assessments in their curriculum. The low priority given to these assessments resulted from the difficulty they met in reaching consensus among themselves about adjustments in teaching. Additionally, alternative assessments made up only a small percentage of students' final grades. They said that they lacked knowledge and confidence in alternative assessment. Indeed, the findings from the study suggested that the summative purposes of assessment in their context influenced their perceptions and practice of assessment.

Unlike the studies reviewed above, Davis and Neitzel's (2011) study examined teachers' perceptions of assessment in terms of the audiences of assessment information. The authors conducted structured observations and semistructured interviews with 15 teachers teaching reading, writing, mathematics and/or science in Grades 4 to 7 in an upper-elementary school and a middle school in the south-eastern United States. Each teacher participated for at least one semester in the

period of three years of the study. The study identified 10 purposes of assessment, which were categorised in terms of audiences: teachers, students, parents, and managers. For teachers, first, assessment informed themselves about their instruction. Second, assessment helped identify students who needed additional support from the teacher or their peers. Third, through assessment, the teachers could judge their students' attainment or understanding. Finally, teachers carried out assessment to gauge their students' investment in classroom activities. For students, assessment served three purposes. The first purpose was making students accountable for their learning through grading to punish or reward them for their academic efforts. Second, assessment was part of instruction because it provided opportunities for students to demonstrate and expand their understanding. Finally, through assessment, students obtained feedback from teachers and peers. Assessment also served external audiences. Assessment outcomes were reported to parents to inform them about their children's learning outcomes. Assessment results were also used by managers to make teachers accountable for their teaching and prepare their students for state-mandated tests. Davis and Neitzel's study showed that the teachers understood that assessment had different purposes for different audiences.

The findings from these studies indicated that teachers in different contexts held multiple perceptions of the purposes of assessment and these perceptions were context-dependent. In general, teachers saw assessment as serving both formative and summative purposes. However, teachers in contexts where assessment was high-stakes (e.g., Hong Kong) placed more emphasis on summative purposes, whilst teachers in contexts where assessment was not high-stakes (e.g., New Zealand) put more weight on formative purposes. In addition to the studies which investigated teachers' perceptions of the purposes of assessment in general, some studies examined teachers' perceptions of assessment of learning.

2.4.2.1.2. Teachers' perceptions of assessment of *learning.*

Some researchers have been interested in teachers' perceptions of summative tests. In a study in Germany, Maier (2009) examined the attitudes towards state-mandated tests expressed by teachers of different school types and school subjects. Maier also examined their thinking about the pedagogical relevance of performance data for their classroom and factors influencing their views about the applicability of tests. Maier (2009) analysed responses to questionnaires completed by 307 Grade 6 teachers in 256 secondary schools in Baden-Württemberg, a state in Germany. State-mandated tests in this context were intended for school self-evaluation and were not high-stakes to students. There were three school types in the study. Grammar schools for Grades 5 to 12 prepared students for university education. They focused on foreign languages, mathematics and science.

Intermediate secondary schools for Grades 5 to 10 prepared students for non-college jobs in administration and services industries. General secondary schools were schools where students studied for five years, preparing for apprenticeships and vocational education.

The results of the study indicated that the teachers from different school types viewed state-mandated tests differently. Specifically, the teachers in the general secondary schools perceived state-mandated tests more positively than those in the grammar and intermediate schools. Also, the subjects that were tested influenced the degree of teachers' acceptance of state-mandated tests. As for the relevance of the mandated test data for instructional improvement, some teachers in the general secondary schools and a few from the other types of schools thought that the data from state-mandated tests served the purpose of improving teaching. The results of the study also indicated that class size and school location affected the teachers' views about the instructional value of state-mandated tests. What is still lacking in this study was how the teachers in the different schools types used the data from state-mandated tests to improve their teaching and students' learning. While the teachers in this study appreciated the formative purpose of large-scale tests, teachers in other contexts may not do so. This was examined in the next study.

Leighton, Gokiert, Cor, and Heffernan (2010) investigated whether or not teachers believed that classroom and large-scale tests were designed to (1) provide information about students' learning processes, (2) influence meaningful student learning, and (3) elicit learning or testing strategies for successful test performance. They asked 272 Grade 7 to 12 teachers from a mid-sized metropolitan area in Canada to respond to a 46-item questionnaire. The teachers believed that classroom tests provided more information about students' learning processes than large-scale tests. They also believed that classroom tests were more likely to influence meaningful learning and required more learning strategies than test-taking strategies. However, the researchers concluded that the teachers in the study held incorrect perceptions about the real value of classroom assessment. The authors cited Stiggins (1991) and Black and Wiliam (1998), who said that classroom assessments did not have more cognitive diagnostic value than large-scale tests and that there was little evidence to believe that classroom assessments were designed to inform student learning processes, to lead to meaningful learning, or to require learning strategies rather than test-taking strategies. The findings in this study suggested that teachers may attribute good values to classroom assessment, even though it does not have such values.

These studies about teachers' perceptions of assessment of learning suggest that teachers do not always have negative beliefs about state-mandated tests. Moreover, they may lack knowledge about assessment and thus attribute to classroom assessments the values that they may not have.

Apart from the interest in teachers' perceptions of summative assessment, researchers have also been interested in teachers' perceptions of assessment for learning.

2.4.2.1.3. Teachers' perceptions of assessment for learning.

Many researchers have carried out studies about teachers' perceptions of assessment for learning. A study of this type was carried out by Hargreaves (2005). Hargreaves asked 83 teachers and principals in primary and secondary schools in the UK to independently write down what assessment for learning meant to them. The teachers believed that assessment for learning (1) helped monitor students' performance against targets or objectives, (2) helped inform next steps in teaching and learning (i.e., to work out what action needed to be taken next), (3) helped teachers give feedback for learning improvement, (4) helped teachers learn about their students' learning, (5) helped students take some control of their own learning and assessment, and (6) helped turn assessment into "a learning event" (Hargreaves, 2005, p. 217). The findings of this study showed that the teachers had very good knowledge and positive perceptions of assessment for learning. However, the study did not investigate how the teachers implemented assessment and how they used assessment data for these purposes. Teachers' perceptions and practices of formative assessment may depend on their particular contexts. This was investigated in the next study.

Remesal (2007) examined why formative practices were more often used in the study's sample of schools in Spain. Remesal interviewed 30 primary teachers and 20 mathematics teachers from secondary schools in Barcelona. Remesal also used the teachers' classroom assessment materials and products and students' classroom assessment results as data in the study. The teachers in the study assessed their students periodically and systematically, using written exams and daily observation as the main assessment techniques. They regarded students' errors as an important part of the learning process. The teachers thought that it was their responsibility to understand students' errors so that they could lead students through a reflection process about their individual errors. The teachers believed that students could only advance in their learning when they recognised their mistakes. In addition, the teachers thought that assessment helped them monitor their teaching and students' learning, evaluate students' progress with respect to the learning goals, and reflect on both their own and their students' actions. However, the teachers also assumed assessment to be "an instrument of social control, a means to certify the students' final results and, thus, it is seen as a way of exposing to the public the teacher's professional activity" (p. 31). In general, both the primary and secondary teachers put more weight on summative purposes of assessment than formative purposes. However, the primary school teachers put more weight on formative purposes, whereas the secondary teachers put more weight on summative purposes. The researcher argued

that the perceptions held by the primary and secondary teachers were distributed differently because secondary education in Spain emphasised the accountability purpose of assessment while the same emphasis did not apply in primary education.

The findings in this study support those in other studies (e.g., G. T. L. Brown, Kennedy, et al., 2009; G. T. L. Brown & Lake, 2006; Cheng, 2006): teachers' perceptions of the purposes of assessment are multiple and influenced by the prioritised purposes of assessment in their contexts. What is not clear in the study is what types of assessment the teachers used for the summative purposes, what types of assessment they used for the formative purposes, and how they used these types of assessment to fulfil these purposes.

In another study, Chan (2007) examined elementary EFL school teachers' beliefs and practices of multiple assessment. In this study, multiple assessment referred to the multiple purposes (e.g., formative and summative), multiple content (e.g., vocabulary, sentence patterns, songs and rhymes as well as the skills of listening, speaking, reading, and writing), multiple tools (e.g., traditional paper-and-pencil tests or use of multiple media such as computer, tape recorder, or video recorder), and multiple testers (e.g., teachers, peers, and self).

Five hundred and twenty elementary EFL teachers in Northern Taiwan were surveyed. Chan (2007) found that these teachers had a very good understanding of multiple assessments. They reported guiding their students to collect material for portfolio assessment. They also set criteria to assess students' language performance. They did not agree that traditional tests were more effective than alternative assessments in assessing reading, writing, speaking, and listening. They believed that multiple assessment led to more focused instruction and increased reflection on teaching practices. They believed that multiple assessment informed them about students' learning difficulties and helped them plan more engaging classroom activities (Chan, 2007). While the teachers in this study appreciated alternative assessments, the Hong Kong teachers in Cheng's (2006) study valued tests and examinations. This may have been because the teachers in Cheng's study lacked knowledge about alternative assessments.

While the above studies investigated teachers' perceptions of assessment, Inbar-Lourie and Donitsa-Schmidt's (2009) study examined the factors underlying Israeli EFL teachers' perceptions and usage of alternative assessments. The researchers analysed the data collected from a questionnaire responded by 113 EFL teachers in elementary, junior-high, and high schools in Israel. It is worth noting that in Israel, both national standardised tests and alternative assessments were used, and teachers were trained in alternative assessment (72% of the teachers in this study had taken at least one course in language testing).

The EFL teachers in Inbar-Lourie and Donitsa-Schmidt's (2009) study believed that alternative assessment was an ongoing, collaborative, and integral part of learning. It helped improve learning and catered to the needs of all students. The teachers reported that there were many obstacles to using alternative assessment in the classroom. These included the lack of time, resources, and teachers' knowledge and skills in designing and implementing alternative assessment; the gap between parents' expectations and the information gained from alternative assessment; and the mismatch between the traditional report card and the detailed qualitative feedback that students received through alternative assessment. Additionally, the teachers were uncertain about the relationship between alternative and external standardised assessment. They believed that all assessments, including alternative assessments, were used for the purpose of monitoring and surveillance. They also believed that external standardised tests hindered the utilization of alternative assessment. The teachers also believed that alternative assessments had low validity and reliability because these assessments focused on the form rather than on the content of students' work. The teachers were unable to ascertain whether students' assessment outcomes really reflected their genuine work. These findings suggested that the teachers could see both the formative and summative purposes of alternative assessment. However, they tended to see the summative purposes more clearly than the formative purposes when they thought that all assessments were used for the purposes of monitoring and surveillance. Like the previously reviewed studies, this study did not investigate how the teachers negotiated the multiple purposes of assessment in their context.

A conclusion that can be drawn from the studies reviewed in this sub-section is that teachers hold multiple perceptions of the purposes of assessment and that their perceptions depend on their specific context. However, it is insufficient just to understand the perceptions of assessment held by the teachers in these contexts in order to understand contexts of assessment. Teachers' implementation of assessment in their classrooms is also an important part of their context of assessment. This is the reason for the review of studies about teachers' assessment practices in the next sub-section.

2.4.2.2. Teachers' practices of assessment.

Studies about teachers' assessment practices have focused on identifying the types of assessment that teachers use in their classrooms. A few studies have investigated more specific areas of assessment such as how EFL teachers assessed students' foreign language performance (Butler, 2009) and the types of materials that teachers of English developed to assess their students (Kahn, 2000).

McMillan (2001) investigated secondary teachers' assessment and grading by using a questionnaire which was completed by1,483 Grade 6 to12 teachers of different subjects in 53 schools in Virginia in the USA. The study found that most of the teachers designed assessments themselves, and relatively few relied on published assessments. The teachers in the lower grades used published tests more often, while teachers in the higher grades used teacher-developed assessments more often. The teachers used these types of assessment: constructed-response assessments (e.g., essays), performance-based assessments, projects, teacher-made assessments, quizzes, objective assessments, and major examinations. Essay-type questions were more often used than objective tests, and student projects and performance assessments were used by many of the teachers. In general, the teachers of different subjects preferred different types of assessment. Moreover, grade level and students' ability level also affected the types of assessment used. The teachers prepared their students for high-stakes tests by using the types of assessment that were included in the tests. In terms of tested skills, the teachers used assessments that measured students' understanding most often, and they also used assessments that measured students' reasoning and application. Assessments that measured students' recall of knowledge were used the least. The teachers' focus on these thinking skills was affected by students' ability level. The teachers used many different types of assessment and the types they chose to assess their students depended on the subject they taught, the grade levels, student ability levels, and on whether they were included in high-stakes tests. One gap in knowledge that this study leaves is how teachers implement assessment in their classrooms and how they negotiate multiple purposes of assessment. The study also suggests that investigations into assessment need to take account of the subject that the teachers teach as part of the context of assessment. Similar to McMillan's study, another study examined the types of assessment the teachers used in an EFL context.

Gattullo (2000) examined how formative assessment was interpreted and implemented by EFL teachers working with students in the final years of primary school in Italy. The data were gathered through observations of four teachers in three schools. The study focused on nine actions: questioning or eliciting, correcting or making counter-suggestions, judging, rewarding, observing processes, examining products, clarifying, setting task criteria, and metacognitive questioning. The study found that the three most used actions were questioning to elicit student understanding, correcting or making counter-suggestions about students' errors, and judging students' work. Metacognitive questioning, observing processes, and examining products were rarely used. The teachers in the study asked questions for two purposes: to revise the content of the lesson and to establish rules for interaction between the teacher and the student(s). However, the teachers did not use metacognitive questions. The study also found that the teachers were often unable to make

productive use of information they collected for formative purposes. For instance, they did not take advantage of students' responses to their questions to understand their students' learning process. This study examined the types of assessment that the EFL teachers used in their classrooms, while the next study examined assessment implemented by teachers of social studies.

Yildirim (2004) examined assessment implemented by teachers in social studies courses in Turkey. Data were collected using a questionnaire and structured interviews. The questionnaire was completed by 531 teachers in 81 high schools in 33 provinces in Turkey. Sixteen interviews were conducted. The results of the study found that the school administrators did not encourage the teachers to use multiple-choice tests because they lacked the expertise in designing and using them. Their administrators also did not encourage them to use essay tests. Oral tests were used just occasionally. The teachers mainly used short-answer tests, even though their students needed to be familiar with multiple-choice tests for the national university entrance examinations. The findings of the study suggested that teachers' assessment practices were influenced by their administrators. The study also suggested that further studies need to investigate the perceptions of both school administrators and teachers to examine how both groups interact and influence each other in the same context.

The above studies focused on identifying the types of assessment that teachers used in their classrooms. The next studies focused on how teachers assessed their students.

Butler (2009) examined how South Korean EFL teachers observed and assessed elementary school students' foreign language performance in classrooms and how such assessments varied among teachers. Butler (2009) asked 26 elementary school teachers (11 teachers taught English only, and 15 teachers taught English and other subjects) and 23 English language teachers in secondary schools in South Korea to watch the videotapes of two group activities undertaken by Grade 6 students in their English language lessons and individually assess the students' performance. After the teachers had finished their evaluation of the students' performance, they discussed in small groups the ways in which they had evaluated each student and the criteria they had used for their evaluation. The discussions were recorded and transcribed for use as a source of data for the study.

Both the elementary and secondary teachers paid attention to some common traits when they evaluated the students' performance: fluency in speaking, confidence when talking, listening comprehension, motivation, and speaking accuracy. However, the teachers in the study, both within and across groups, were very different in their holistic evaluations of student performance. This was because the teachers interpreted the students' traits and made judgements about them in different ways. While the elementary teachers avoided setting any criteria, the secondary teachers relied on

criteria when judging the students' performance. While some teachers emphasised accuracy in speaking, others gave priority to fluency. The secondary teachers emphasised students' affective aspects such as confidence more than the elementary teachers, who placed more attention on the linguistic aspects of the students' performance. It is clear that expectations of the elementary and secondary teachers for the elementary students' performance were different.

According to Butler (2009), the differences between the elementary teachers' and secondary teachers' practices of assessing elementary school students' English performance were attributable to their respective teaching and assessment contexts. In South Korea, the change from the grammar-translation language teaching approach to the communicative language teaching approach called for an emphasis on affective aspects such as confidence in talking and motivation. The call also brought about the avoidance of measurement-oriented assessment and competition in the English curriculum among elementary students. The next study also found that the teachers' assessment practices were influenced by their contexts.

To understand the perceptions of teachers of English, Kahn (2000) examined the assessment materials designed by a group of teachers teaching Grade 10 students of English in a large suburban high school in the Midwest of the USA. The materials included all the major tests, the final examination, and other materials used for assessment purposes in the first semester of a school year. In addition to the materials, notes from informal conversations and discussions with the teachers were a source of data for the study.

Kahn's (2000) study showed that the assessment materials included a mixture of approaches. Some materials required students to construct meanings and interpret texts for themselves. Students' ability to apply concepts to new problems or situations was assessed in materials that required extended written or spoken responses. In this approach, both the content and form were valued. Written or spoken responses were the preferred formats of the materials used to assess writing and speaking. However, memory and recall were the main foci in these assessment materials. Specifically, about 65% of the total points for the semester involved multiple-choice tests. The preferred formats for assessing literature, listening, and especially grammar and vocabulary, were multiple-choice tests which assessed memory and recall. The reason for prioritising multiple-choice tests was that this format was used in final major tests and examinations, so tests with this format were thought to help maintain students' attention, cooperation, and classroom control. The results of the study suggested that teachers' practices of assessment was influenced by high-stakes tests in this contexts.

The studies reviewed in this section show that since 2000, researchers have been interested in teachers' perceptions of the purposes of assessment and the types of assessment used in

classrooms. There are still some gaps in the body of knowledge about the contexts of assessment, teachers' perceptions of their contexts of assessment, and teachers' implementation of assessment in their classrooms. Specifically, further research is needed to know more about assessment in particular contexts such as classrooms and in different educational jurisdictions. In order to know about a context of assessment, it is insufficient just to examine what teachers think about the purposes of assessment and what assessment types they use. It is essential to know what teachers think about their context of assessment, how they implement assessment in their classrooms, and how they negotiate the multiple purposes of assessment. These are the concerns of the present study.

2.4.3. Students' perceptions of assessment.

Research studies about students' perceptions of assessment have focused on students' perceptions of the purposes of assessment (G. T. L. Brown & Hirschfeld, 2007, 2008; G. T. L. Brown, Irving, et al., 2009), classroom assessment (Brookhart & Bronowicz, 2003; Moni et al., 2002; Yildirim, 2004), and high-stakes assessment (Triplett & Barksdale, 2005).

2.4.3.1. Students' perceptions of the purposes of assessment.

Brown and Hirschfeld (2007) investigated secondary students' perceptions of assessment and the relationship between these perceptions and their achievement in standardised national assessments of mathematics. They analysed 1191 questionnaires from Grade 9 to 12 students in secondary schools in New Zealand. The questionnaires comprised 49 items related to four main purposes of assessment: assessment made students accountable, assessment improved teaching and learning, assessment was negative or bad, and assessment provided a useful description of performance. The students were asked to show their agreement with the item on the questionnaires.

The students perceived that assessment made students and schools accountable. They also thought that assessment was useful because it helped improve teaching and learning. Therefore, they disagreed that assessment was negative or bad.

These two researchers conducted a similar study to investigate students' perceptions of assessment and how these perceptions linked to their achievement on standardised national assessments of literacy (G. T. L. Brown & Hirschfeld, 2008). In this study, the researchers elicited students' perceptions of the purposes of assessment by using questionnaires with 29 items which were responded to by 3469 students in 58 secondary schools in New Zealand. The students were asked to indicate to what extent they agreed or disagreed with four purposes of assessment: assessment made students accountable, assessment made schools accountable, assessment was fun,

and assessment was ignored. The students perceived that assessment made themselves and schools accountable and assessment was fun. They did not agree that assessment was ignored.

In 2009, Brown, Irving, et al. (2009) examined how students' perceptions of assessment were related to each other, how students defined assessment, and how students' perceptions were related to their definitions of assessment. These researchers analysed questionnaires responded to by 705 Grade 9 and 10 students in 31 secondary schools across New Zealand. These questionnaires consisted of 45 items which asked the students to indicate the degree of their agreement with each item. In addition, the students were asked to indicate which of the 12 given assessment practices they associated with the word assessment. These assessment practices were regarded to be common in New Zealand schools.

The students' perceptions of the purposes of assessment were hierarchical. Specifically, they mostly agreed that assessment was used by teachers and students to improve student learning and moderately agreed that assessment made students and schools accountable. They slightly agreed that assessment was personally enjoyable and affectively beneficial for the classroom environment (i.e., assessment created an environment where students worked together and helped each other). They believed that assessment that did not help students and teachers improve student learning was irrelevant.

The students associated the term assessment mainly with test-like or teacher-controlled assessment practices (e.g., examinations, essays), thus they did not pay much attention to interactive-informal assessment practices such as self-assessment, peer-assessment, oral classroom questions, teachers' observations of student learning in the classroom, portfolios, and teacher-student conferences.

The researchers found that when students believed that teachers used assessment to improve student learning, they believed that teacher-controlled assessment practices should be used and they did not think that assessment was personally enjoyable. Students believed that when interactive-informal assessment practices were used, assessment affectively benefited the classroom environment but it was ignored.

The studies carried out by Brown and colleagues reviewed above generated several findings about students' perceptions of assessment. However, the findings of these studies mainly related to students' perceptions of the purposes of assessment. Additionally, these studies were carried out in the same context (i.e., New Zealand) where assessment in secondary schools was low stakes and test results did not have substantial consequences for either the students or the schools (G. T. L. Brown & Hirschfeld, 2008). Studies of students' perceptions of assessment in other countries should also be considered.

2.4.3.2. Students' perceptions of classroom assessment.

High school students' perceptions of literacy assessment in English were examined by Moni, et al. (2002). Specifically, these researchers examined what the students thought, felt, and valued about literacy assessment and whether or not their thoughts, feelings, and values changed throughout their first year in high school. They observed the students of two Grade 8 English teachers in two high schools in Queensland, Australia and individually interviewed 18 Grade 8 students in these two classrooms.

A number of students believed that the main use of assessment in high school was to obtain marks for report cards, while other students thought that assessment informed teachers about students' learning. Several students believed that assessment helped them improve their learning and helped teachers monitor students' performance. Some students believed that assessment data were important for job references. Many students were not satisfied with the teachers' red crosses as feedback, and some students believed that marks were not accurate measures of their learning outcomes because marks depended on teachers' personal feelings about individual students. A number of students believed that their teachers paid attention to surface matters (e.g., spelling, punctuation, and neatness), attitudes, and personality.

At the beginning of the first school year in high school, many students were not familiar with task sheets as assessment tools because task sheets were not used when they were in primary school. Gradually, many students valued detailed explanations and instructions in task sheets. A number of students found writing responses to poems difficult because of the unfamiliar aspects of the task. Most students perceived autobiographical writing was easy because this task was familiar to them and they wrote about themselves. A number of students did not like oral assessment because they felt embarrassed when responding to oral questions.

The students perceived that in high school, they undertook assessment more frequently than in primary school and the tasks had a higher volume of work.

Yildirim (2004) investigated how students perceived assessments in social-studies courses in high schools in Turkey and how these assessments influenced teaching and learning. The findings reviewed in this section were related to the students' perceptions, and the findings related to the teachers' perceptions were previously reviewed in the review of teachers' assessment practices.

The researcher used questionnaires and focus group discussions to collect data for the study. Eight hundred and eighteen students from 81 high schools in 33 provinces in seven regions in Turkey responded to the questionnaires. The questionnaires comprised both closed-ended and open-ended questions about the types of assessment techniques used, their effects on teaching and

learning, and recommendations for improving student assessment. Eighty-five students participated in focus group discussions; each group included five to six students. Interview questions were aligned with the questionnaires.

The majority of students did not think that the way they were assessed was adequate for measuring their performance. They perceived that sometimes too many topics were tested, and when this occurred, they did not know how to prepare for a test. They stated that testing too many topics in a test made them stressed and confused.

The students thought that multiple-choice tests helped them prepare for the national university entrance examination. Moreover, they liked multiple-choice tests because they did not have to formulate responses and they felt less nervous and stressed. Although they found short-answer tests – the main format used in their schools – boring, they preferred them to other formats because they were familiar to them and they knew how to prepare for these tests. They believed that these tests assess only factual knowledge (e.g., dates, definitions, and descriptions). They suggested that teachers should assess students' interpretation and application of what they knew because these skills were more important in the long run. The students suggested that instead of including too much content in a test, teachers should give them quizzes and tests after each major unit because this would reduce what they had to learn for each test.

Although essay tests were given only occasionally, the students disliked these tests. They found it difficult to know exactly what essay questions asked and how to respond to them. Some students argued that essay tests were not objective and fair because different criteria may be used to judge students' essays.

A number of students did not like oral tests because they felt nervous when responding in front of peers. They preferred written tests to oral tests because they had some time to think when responding to written tests. Some students suggested that their total classroom performance should be considered when teachers made judgements about their learning rather than only using oral tests. They believed that this way of judging their achievement encouraged them to participate more actively in classroom activities and do assigned readings regularly.

Similarly, Brookhart and Bronowicz (2003) investigated students' perceptions of how students viewed a particular classroom assessment and its importance. The authors also investigated the students' abilities to do the assessment, their reasons for doing it, and the effort they expended. The authors interviewed 161 students (63 elementary and middle school students and 98 high school students in the USA). In addition, they observed 36 assessment events in the classrooms, which ranged from traditional paper-and-pencil tests to group presentations.

Students were asked about the task characteristics (interest and importance), difficulty, self-efficacy, and goal orientations (whether the task itself was important or whether external approval was important to the student). Interview questions were specific to each classroom assessment.

Elementary and middle school students perceived that their teachers expected them to pay attention to both the content and mechanics of assessment (i.e., formats of tests and procedures for doing assignments). They believed that their teachers expected them to do well or to give correct answers to all questions. High schools students perceived that their teachers expected them to study and revise lessons for tests. However, the students rarely mentioned the specific content and skills that they were expected to learn.

High school students believed that their teachers expected them to find sources of reference when they did research projects. They were concerned about finding information for their projects and perceived this to be difficult. For group work, they perceived that their teachers expected them to learn how to work with each other and how to divide up tasks among the group members. Some students felt confident because they worked with peers. However, a number of students found it difficult to work in a group because the group members may not agree with each other.

Elementary and middle school students mentioned two reasons for them to assume that an assignment was important. First, the assignment was either related to immediate goals (e.g., they needed to know how to do a similar task in the next grade) or future goals (e.g., the assignment was related to their future job). High school students perceived that an assignment was important because learning the content and skills was important for college learning or for a specific career. A number of high school students perceived assignments to be important because they were related to real life (e.g., knowing how the human body functions, knowing about the past). Other high school students perceived an assessment task to be important because the knowledge and skills learned were important in and of themselves (e.g., the knowledge and skills learned were important for leading an informed life).

Elementary and middle school students were interested in an assignment when they perceived that it was interesting to them, it was useful for them, or it was within their ability. High school students found an assignment interesting when they perceived that it was important in some way (e.g., it could be used in a later assignment). A few high school students perceived an assignment to be important even though they did not find it interesting. These students perceived that the assignment was important because they may have to do similar assignments later.

Citing Ames's (1992) theory of learning, Brookhart and Bronowicz (2003) argued that students with a mastery goal orientation wanted to learn because they wanted to master the material or skills that they found worthwhile, while students with a performance goal orientation wanted to

learn because they wanted other people to think that they were clever, competent, or talented. The study found that high school students usually did not care about others' performance on an assessment, and they usually did not care how others thought about their own performance on an assessment. A few students thought that others should not care about their performance on an assessment. Several students reported that they cared how others thought about their performance when they wanted others to learn from their project or report, when they wanted to learn from others' projects or reports, or when they wanted approval from peers. Some students cared about others' performance because of sympathy or altruism.

Based on these findings, the authors concluded that the students first and foremost mentioned their own needs and interests, no matter what the assessment was. There was a difference between elementary and middle school students and high school students in perceptions of teachers' expectations. Grade 3 and Grade 5 students perceived that their teachers were concerned about the quality of student performance (i.e., to do well in tests, to complete all the questions), while high school students believed that their teachers expected them to pay more attention to the learning process.

The review of the studies in this sub-section indicated that students perceived that assessment had various purposes. In addition, students did not believe that assessment results accurately reflected their performance. They found a type of assessment difficult if it was new to them. Therefore, they preferred a type of assessment that was familiar to them, even though they perceived that this type of assessment was not the best type of assessment in terms of assessing their performance, helping them develop skills that they perceived to be worthwhile, or preparing them for major examinations. The findings in the studies reviewed in this sub-section also indicated that students at different school levels held different perceptions of what teachers expected from them, what assignments were difficult, and why an assignment was interesting or important. These findings suggested that research needs to examine what perceptions students in a particular context of assessment hold about the purposes of assessment. In addition, research needs to investigate what students perceive their teachers expect from them in assessment, what types of assessment they prefer, and why they prefer these types.

2.4.3.3. Students' perceptions of high-stakes assessment.

Elementary students' perceptions of high-stakes testing were examined by Triplett and Barksdale (2005). One day after 225 Grade 3 to 6 students completed their high-stakes tests, the researchers asked each to draw a picture about his or her recent testing experience and write to explain the picture. These students were from five schools reflecting racial and socioeconomic

diversity. Half of them were from a rural area in a mid-Atlantic state and the other half were from an urban area in a southern state of the USA. These two states were chosen because one state had an established history of high-stakes accountability, including using test scores for promotion and retention and ranking schools and teachers in local newspapers. The other state only began to use accountability measures in response to the No Child Left Behind Act.

The students' drawings and written descriptions showed that high-stakes testing created negative emotions (e.g., nervous, frustrated, confused, and angry). Of the negative emotions demonstrated or written about, nervous was most frequently mentioned, and this feeling was related to not being able to work out the answers, not having enough time to finish, and not passing the test. Another common emotion among the students was anger over the length and difficulty of the test, not being allowed to talk to peers, and possible consequences of failure. A few students expressed positive emotions such as "glad", "happy", or "liked". However, these emotions were related to things beyond the test (e.g., being given chewing gum or ice-cream after the test). Very few students used "easy" to refer to the test experience.

A number of students' drawings and written descriptions referred to content areas, and most of them indicated that maths tests were most difficult for them. Only a few students' drawings and written descriptions referred to teachers, and these students described their teachers as monitors (monitoring and controlling students in testing settings), coaches (providing motivational comments), comforters (giving chewing gum), and uninterested observers (sitting in the classroom without any interaction with students).

A number of students used metaphors (e.g., black clouds) to express their emotions and feelings when taking tests. In addition, they used fire, flames, and burning to express their anger and powerlessness. Many students perceived that the state wanted to assess how clever they were and that tests were used to make decisions about promotion and retention. Some students suggested that people should protest against high-stakes testing. Several students described the consequences of passing and failure on their immediate and near future (e.g., going to high school, going to college, or becoming a Burger King driver). Over half of the students referred to accoutrements of testing in their drawings and written descriptions (e.g., calculators, the date on answer sheets, test booklets, and clocks). A great number of students drew a child sitting alone at a desk without anyone else in their pictures. This indicated that the students felt isolated in testing settings. The researchers concluded that the perceived consequences of high-stakes testing made students feel nervous, angry, and powerless.

The findings of the studies reviewed in this section indicated that students perceived that assessment had multiple purposes. Students had various positive expectations of assessment.

However, they held many negative perceptions of assessment, especially in contexts where assessment was high-stakes (Triplett & Barksdale, 2005).

Most of the studies reviewed in this section were carried out in Western countries, including New Zealand (G. T. L. Brown & Hirschfeld, 2007, 2008; G. T. L. Brown, Irving, et al., 2009), the USA (Brookhart & Bronowicz, 2003; Triplett & Barksdale, 2005), and Australia (Moni et al., 2002). Assessment in high schools in Vietnam is administered to generate marks for ranking students and to make decisions about promotion and retention. Therefore, high schools students in Vietnam may hold perceptions which are different from those found in studies conducted in Western contexts. In addition, previous studies have not examined how high school students negotiate the multiple purposes of assessment. The present study will fill in these gaps.

2.5. Summary

This chapter has argued that assessment *for* learning, which is associated with a culture of assessment, is grounded in the sociocultural constructivist theory of learning and that assessment *for* learning is typical of a culture of assessment. Assessment *of* learning is typical of a culture of testing. The chapter also provided a review of the literature on principals', teachers', and students' perceptions of assessment and teachers' assessment practices. The chapter argued that research into principals', teachers', and students' perceptions of assessment needs to look carefully at the contexts from which these perceptions emerge. Specifically, the Vietnamese EFL high school context will be the setting for the present study, and the perceptions of assessment of the various groups in this context will be examined. In addition, the study will explore assessment practices used in this setting and how these stakeholders negotiate the multiple purposes of assessment. The next chapter addresses how data were collected and analysed to answer the research questions.

Chapter 3. Method

This chapter first presents the research approach chosen for this study, information about the trustworthiness of the study, the design of the study, the research sites, and participants. The chapter also addresses the instruments used to collect data, procedures for data collection, and how the collected data were stored and analysed. The procedures for ethical clearance and participant recruitment are presented in the last section of the chapter.

3.1. Qualitative Approach

The social researcher's decisions about the research approach reflect his or her ontological and epistemological assumptions (Denzin & Lincoln, 2003; Neuman, 2014). Ontology is related to "the fundamental nature of reality" or "the nature of being" and answers the question "what exists" (Neuman, 2014, p. 94). Therefore, the researcher's ontological assumptions are related to what he or she considers to be reality. Epistemology is defined as "[a]n area of philosophy concerned with the creation of knowledge; focuses on how we know what we know or what are the most valid ways to reach truth" (Neuman, 2014, p. 95). This definition suggests that the researcher's epistemological assumptions are related to the nature of knowledge and how knowledge is acquired.

Different ontological and epistemological assumptions shape different research paradigms. A paradigm is defined as "[a] general organizing framework for theory and research that includes basic assumptions, key issues, models of quality research, and methods for seeking answers" (Neuman, 2014, p. 96). The quantitative approach and the qualitative approach rest on two different paradigms.

The quantitative approach rests on the positivist or realist paradigm (Creswell, 2009; Denzin & Lincoln, 2003; Hittleman & Simon, 2002; Neuman, 2014). Positivists assume that "the 'real world' exists independently of humans and their interpretations of it" and the world "is organized into preexisting categories just waiting for us to discover" (Neuman, 2014, p. 94). Neuman (2014) states that researchers who adopt the quantitative approach try to measure objective facts, and hence they focus on distinct variables. They believe that research should be value-free. Theories, mainly related to causal relationships, are formulated in the form of hypotheses. Data are collected and analysed in order to confirm or disprove hypotheses. Because data in quantitative studies are statistically analysed, large samples are essential for reliable claims and generalisations. In quantitative studies, data are presented in numbers and/or percentages in tables or charts. Quantitative researchers are expected to be objective in collecting and analysing data and in

presenting the results of their data analysis. Therefore, they are expected to be detached from their participants.

The qualitative approach is based on the constructivist, interpretivist, post-positivist, naturalist, or nominalist paradigm (Creswell, 2012; Lincoln & Guba, 2003; Neuman, 2014). The constructivist believes that there are multiple realities (Stake, 2006) because people always see things "through a lens or scheme of interpretations and inner subjectivity" (Neuman, 2014, p. 94). That is, constructivists assume that knowledge depends on individuals' interpretations, values, and contexts (Creswell, 2009; Greene, Kreider, & Mayer, 2005). Therefore, the qualitative researcher's task is to "understand, interpret, seek meaning, describe, illuminate and theorise" (Herbert & Higgs, 2004, p. 64) or "understand and interpret the world in terms of its actors" (L. Cohen, Manion, & Morrison, 2000, p. 181).

Neuman (2014) says that qualitative researchers try to derive insights into social realities and cultural meanings in context. Therefore, their studies focus on interactive processes and events. Data are in the form of words from documents, observations, and transcripts. Data can also include images and artifacts. Theories may or may not be available before data are collected and analysed. Preferably, theories are grounded in data. That is, theories are formulated through data analysis in terms of categories, themes, and models. In qualitative research, subjectivity, defined as a subject's "particular perspective, feelings, beliefs, and desires" (Sol, 2005, p. 900), is explicitly acknowledged because researchers are involved in the interactive processes and events that they study. Unlike quantitative studies, qualitative studies do not make generalisations from large samples. Instead, qualitative researchers choose a few subjects or cases that have the knowledge and experience they require and the ability to reflect (Andrade, 2009). They also choose participants who are articulate, have the time to be interviewed, and are willing to participate in their studies (Andrade, 2009) because researchers try "to understand a few cases in depth rather than represent an entire population" (O'Reilly, 2005, p. 225).

Therefore, the qualitative approach was suited to the present study. In the present study, the researcher tried to understand and interpret the principals', EFL teachers', and students' perceptions and practices of assessment in their schools (i.e., specific contexts) and how they interpreted policy-makers' and other stakeholders' expectations about assessment in English instruction at the two high schools and how they interpreted their own activities in these contexts.

3.2. Trustworthiness of the Study

Qualitative research studies are evaluated according to their trustworthiness (Angrosino, 2007; Gay & Airasian, 2000; Graneheim & Lundman, 2004; Teddlie & Tashakkori, 2009).

Trustworthiness is established when findings in a study reflect the meanings made by the participants (Lincoln & Guba, 1985). Therefore, qualitative researchers try to increase the confidence that their findings represent the meanings presented by their research participants (Lietz, Langer, & Furman, 2006). They give "priority to the meanings of the participants over the perspective of the researcher" (Lietz et al., 2006, p. 443) and try to "describe qualitative data in a way that is credible" (Lietz et al., 2006, p. 444).

Threats to trustworthiness include participants' reactivity (i.e., participants' adjustment of their activities when they are aware of being observed (Padgett, 1998) and the researcher's bias (Lietz et al., 2006; Padgett, 1998). Multiple strategies have been suggested for minimising these threats, including reflexivity, peer debriefing, using an audit trail, triangulation, member checking, and prolonged engagement (Creswell, 2003; Golafshani, 2003; Graneheim & Lundman, 2004; Horsburgh, 2003; Johnson & Waterfield, 2004; Li, 2004; Lincoln & Guba, 1985; Mauther & Doucet, 2003; Padgett, 1998).

Reflexivity is defined as the researcher's acknowledgement that his or her actions and decisions inevitably have an influence on the meaning and context of the phenomenon under investigation (Horsburgh, 2003). The researcher identifies his or her subjectivity and shows how his or her beliefs, experiences, and identity meet those of the participants (MacBeth, 2001). The researcher of this current study acknowledged that his knowledge of assessment in school education in Vietnam may influence his interpretations of the policy documents, the participants' accounts in the interviews, and the classroom observations.

The present researcher was responsible for his study. However, he acknowledged that his supervisors took an important role during the analysis of the data. The regular meetings with the supervisors were sessions for "peer debriefing" because at these meetings the researcher explained how he analysed the data and how he obtained the categories, themes, and theme clusters. The researcher also explained to his supervisors what the categories, themes, and theme clusters meant and how they were related.

The term audit trail refers to the researcher's detailed description of the research procedures that they implement (Johnson & Waterfield, 2004; Padgett, 1998). In this thesis, the researcher kept a record of the procedures that he employed during data collection and recorded in notes how he analysed the data as he progressed.

"Triangulation by observer" (Lietz et al., 2006, p. 451) was also used in the present study. Triangulation by observer is implemented when different researchers look at the data (Lietz et al., 2006). The present researcher and his supervisors independently analysed the principals' interviews. This helped increase the trustworthiness of the findings because commonalities and differences

between the researcher's and his supervisors' analyses were further investigated and opposing perspectives helped increase understanding of the data (Creswell, 2003; Johnson & Waterfield, 2004; Li, 2004). In addition, the researcher's general familiarity with the assessment practices in schools in Vietnam (although not of the assessment practices of the two schools in the study); his immersion in the schools; his interaction with the principals, the EFL teachers, and students in the schools for nearly one semester; the supervisors' unfamiliarity with the assessment practices in Vietnam; and their expertise in qualitative data analysis appeared to complement each other in interpreting the data. Other types of triangulation include data triangulation and method triangulation. These types of triangulation are used to crosscheck results for consistency and to offset any bias in order to reduce the chances of reaching false conclusions (Hammersley, 2008), to reduce uncertainty in data interpretation (Webb, Campbell, Schwartz, & Sechrest, 2000), and to increase confidence and accuracy in the overall conclusions drawn from the study (Spicer, 2004). In data triangulation, the researcher uses various data sources in a study (Spicer, 2004; Teddlie & Tashakkori, 2009) to gain insights into the context and the phenomenon under investigation (Taylor & Bogdan, 1998). Specifically, the data of the present study included policy documents related to assessment in school education; interviews with the principals, EFL teachers, and students; and classroom observations. Method triangulation involves the use of multiple methods to collect the data for a study (Teddlie & Tashakkori, 2009). The methods used to collect data for the present study included semi-structured interviews and classroom observations. The various methods of data collection and different sources of data helped the researcher of the current study triangulate the data.

Member checking, also called respondent validation, allows the participants to review the research findings in order to confirm or challenge the accuracy of the researcher's interpretations of the data that the participants have provided (Creswell, 2003; Horsburgh, 2003; Johnson & Waterfield, 2004; Lincoln & Guba, 1985). This is an important strategy for establishing trustworthiness because it helps reduce the threat of the researcher's bias (Padgett, 1998). However, due to their excessive workload, the participants in the current study refused to read the findings of the study before they were reported in this thesis.

Prolonged engagement refers to the extensive time that the researcher spends with the participants (Lietz et al., 2006). Prolonged engagement is important to the trustworthiness of a study because it helps increase rapport between the researcher and the participants, which makes the participants more open in their interactions with the researcher (Lietz et al., 2006). The researcher of the current study spent more than three months with the participants and built a good rapport with them. In addition, the participants felt secure in sharing their perceptions and experiences when

they were informed that the interviews and classroom observations would serve research purposes only and that the schools' names and the participants' names would not be used.

3.3. Multiple-Case Study Design

Yin (2003) argues that the choice of a research strategy (e.g., experiment, survey, archival analysis, history, or case study) is informed by three conditions. These conditions include "(a) the type of research question posed, (b) the extent of control an investigator has over actual behavioral events, and (c) the degree of focus on contemporary as opposed to historical events" (p. 5).

Yin (2003) suggests that "who," "where," "how much," and "how many" questions are best answered by survey and archival analysis strategies, whilst "how" and "why" questions are better answered by history, experiment, and case study strategies. Yin also suggests that "what" questions can be answered by any research strategy (i.e., experiment, survey, archival analysis, history, or case study) when they require exploratory studies. The proposed study investigated the principals', the EFL teachers', and their students' perceptions of the contexts of assessment in their schools, how classroom assessment was implemented in the EFL classrooms, and how the principals, the EFL teachers, and students negotiated the multiple purposes of assessment in their schools. Therefore, a qualitative multiple-case study was appropriate for answering the research questions.

In terms of the control over and access to actual behavioural events, the aim of this study was not to control the assessment events under investigation. Rather, the researcher relied on the interviews with the principals, EFL teachers, and students as well as direct observations of assessment practices in the classrooms in order to know the principals', the EFL teachers', and the students' perceptions to identify the EFL teachers' assessment practices and the way these stakeholders negotiated the multiple purposes of assessment in their schools.

Case studies include single-case studies and multiple-case studies. Single-case studies are appropriate when researchers want to study critical, extreme or unique, representative or typical, revelatory cases, or when they can study the same single case at two or more different points in time (i.e., a longitudinal case) (Yin, 2003). Researchers choosing to study single cases may face the risk of working with the wrong cases because a critical case may turn out not to be critical when the researcher is working in the field (Yin, 2003). To avoid this risk, multiple-case studies are preferable.

Yin (2003) argues that even with only two cases, the conclusions drawn independently from multiple cases are more powerful than those drawn from a single case. In addition, if common conclusions can be drawn from cases with varied contextual features, the external generalisability of

the findings of the overall case study is increased compared to those from a single-case study. Given these strengths of multiple-case studies, a multiple-case study design was used in this study.

Therefore, this study used a multiple-case embedded design, which encompasses more than one case, with more than one unit of analysis in each case. The multiple-case embedded design was adopted in this study with the assumption that multiple cases (i.e., high schools) and multiple units of analysis (i.e., the principals, EFL teachers, and their students) helped create a more comprehensive picture of the participants' perceptions of the contexts of assessment in their schools; how the EFL teachers implemented assessment in their classrooms; and how these stakeholders negotiated the multiple purposes of assessment in their schools.

The results of the analysis of the interviews with the two principals indicated similar findings. Thus, the hypothesis made by the researcher at the start of the study that school type (i.e., the public school versus the private school) might influence the outcomes of the study was not substantiated. The perceptions of the principals and the school type also did not have an influence on the teachers' perceptions and practices of assessment.

3.4. Research Sites

Two high schools in a city in Central Vietnam were chosen as the research sites. One was a public high school, and the other was a private high school. The public school was established nearly one hundred years ago. The budget for the school's facilities and salaries were provided by the Provincial Committee. The principal was appointed by the DOET, and the teachers were recruited by the DOET. The school administrators and teachers worked according to the regulations and directions of the MOET and the DOET. Because the more able 9th graders in the city had to take a competitive examination in order to enrol at this school, the school administrators and the teachers were under pressure to maintain the school's reputation as one of the best high schools in the province.

The private school had operated for less than 10 years at the time of the data collection. The budget for the facilities, salaries, and expenses of the school came from the shareholders. The school's Management Board appointed the principal, who was also the chairman of the Management Board, and recruited the teachers. Most of the students who had enrolled at this school were less able 9th graders in the city. These students could not enter the public schools in the city because their examination scores were not high enough for entry into these schools. The school administrators and teachers at this school were trying to build their reputation in order to attract and enrol more able students. They used the same textbooks and curriculum that all high schools in the country used, and were controlled by the same regulations related to instruction and assessment.

These high schools were chosen because it was thought that they may have worked under different types of pressure. G. T. L. Brown (2004) and Hill (2009) have also suggested that self-management at the school level influences teachers' perceptions and practices of assessment (G. T. L. Brown, 2004; Hill, 2009). Even though both these schools operated according to the same regulations issued by the MOET and the DOET, it was thought by the researcher that they may have implemented these regulations differently. Therefore, the principals, the EFL teachers, and students in these two schools may perceive and implement assessment differently. Nevertheless, the schools could be considered as representative of public and private high schools in mainstream education at this school level in Vietnam.

3.5. Participants

Qualitative researchers try to look for participants who have the knowledge and experience that researchers require, the ability to reflect, the time to be interviewed, and who are articulate, and willing to be participants in the study (Morse, 1994). The participants in this study were two principals, six EFL teachers, and 72 students in two high schools in Vietnam. In each school, the principal, one EFL teacher from Grade 10, one from Grade 11, one from Grade 12, and 36 students in three classes participated.

There was a reason for choosing EFL teachers in high schools as the participants in this study. Although English is taught in all school levels (i.e., primary/elementary schools for Grades 1 to 5, secondary schools for Grades 6 to 9, and high schools for Grades 10 to 12) in Vietnam, teachers and students in high schools have to prepare for the General Education Diploma Examination (GEDE) and the National University Entrance Examination (NUEE). English is a compulsory subject in the GEDE for all students and in the NUEE for many majors. It was assumed that these examinations had a powerful influence on principals', EFL teachers', and students' perceptions and the practices of assessment.

Personal factors such as training, qualifications, gender, and years of administration or teaching experience of the principals and EFL teachers in the study were not included in the report due to ethical considerations in research. Specifically, the school types and locations were possibly identifiable and details regarding the two principals and six EFL teachers who could potentially have made them identifiable.

3.6. Instruments and Data Collection

Data for the study comprised policy documents related to assessment in school education in Vietnam, two semi-structured interviews with the two principals, six semi-structured interviews

with the six EFL teachers, 12 focus group interviews with 72 students, and 42 classroom observations.

These data sets served various purposes. The results of the analysis of the policy documents provided knowledge about the broader context of assessment in school education in Vietnam. Specifically, policy documents indicated what policy-makers expected from reform in assessment, what purposes of assessment they expected, what methods of assessment they prioritised, and what they valued in assessment. The analysis of the individual interviews with the principals and EFL teachers and the focus group interviews with the students answered the first research question (i.e., the participants' perceptions of the contexts of assessment in their schools) and the third research question (i.e., how they negotiated the multiple purposes of assessment in their contexts). The classroom observations provided the data to answer the second research question (i.e., how the EFL teachers implemented assessment in their classrooms) and the third research question.

3.6.1. Individual semi-structured interview.

Interviewing, defined as "a process in which a researcher and participant engage in a conversation focused on questions related to a research study" (deMarrais, 2004, p. 55), is an appropriate method of data collection when the researcher is interested in "the lived experience of other people and the meaning they make of that experience" (Seidman, 2006, p. 9). Specifically, through interviewing, the researcher tries to find out from participants what she or he cannot directly observe by asking them about their feelings, thoughts, intentions, perspectives, or descriptions of specific experiences (deMarrais, 2004; Patton, 2002). In this study, individual semi-structured interviews were conducted with the principals and the EFL teachers. Individual semi-structured interviews were conducted because they are flexible enough to include both predetermined open-ended questions and those that emerge from the dialogue between the interviewer and the interviewee (Dicicco-Bloom & Crabtree, 2006). The interview protocols were written in the native language of both the researcher and the participants (i.e., Vietnamese), and the individual interviews were conducted in Vietnamese. The individual interviews were conducted throughout one semester, from February to May, 2012.

Each principal participated in one semi-structured interview of around 60 minutes. The principals answered the pre-defined questions written in the *Protocol for the Interviews with the Principals* (Appendix A). These questions elicited the principals' perceptions of the meaning of assessment, their perceptions of assessment in their schools, their expectations of assessment in English language teaching, their perceptions of influences on assessment practices in their schools, their perceptions of the EFL teachers' knowledge and skills in assessment, their support for the EFL

teachers in assessment, and their perceptions of the EFL teachers' autonomy in assessment. In addition to the pre-defined questions, probing questions were asked during the interviews. Probes, defined as "questions or comments that follow up something already asked" (Merriam, 2009, p. 100), were used to ask the principals for more details, clarifications, illustrations, or descriptions of what they said during the interviews.

Each of the six EFL teachers participated in one semi-structured interview before the classroom observations were conducted. Each interview lasted about one hour. The pre-defined questions written in the *Protocol for the Interviews with the EFL Teachers* (Appendix B) elicited the EFL teachers' perceptions of the meaning of assessment. These questions also prompted the teachers to articulate the practices and purposes of assessment in their classrooms and schools. In addition, the teachers were asked to tell how assessment should be implemented and what influenced assessment in their classrooms and schools. The pre-defined questions also asked the teachers about their perceptions of the MOET's, their principals', their students', parents', and their own expectations of assessment. One of the pre-defined questions asked the teachers to tell how they negotiated various stakeholders' expectations of assessment. Probing questions were also used during the interviews when more explanations, illustrations, and descriptions of what they said were deemed necessary.

3.6.2. Focus group interview.

"Focus group discussion" and "focus group interview" can be used interchangeably and are defined as "an interview on a topic with a group of people who have knowledge of the topic" (Merriam, 2009, p. 93). For consistency in terminology, focus group interview(s) was used in the present thesis.

Focus group interviews allow the researcher to access both the individual and collective ideas and stories of the participants (Baartman, Bastiaens, Kirschner, & Vleuten, 2007). Additionally, focus group interviews allow the individual participants to build on other members' opinions and thoughts (Baartman et al., 2007). Therefore, focus groups interviews were appropriate for collecting data about the students' perceptions of the contexts of assessment in their classrooms and schools and how they negotiated the multiple purposes of assessment.

The first 12 students with even numbers on each class list were selected and put in two groups for the focus group interviews. Twelve focus group interviews were conducted with 12 groups of students selected from six classes taught by the six EFL teachers. Each group consisted of six students. Each focus group interview lasted around 60 minutes. The focus group interviews were prompted by the pre-defined questions written in the *Protocol for the Interviews with the Students*

(Appendix D). The students were asked to articulate their understandings of the meaning, practices, and purposes of assessment in their classrooms and schools. They were also asked to express their preferred methods of assessment and the reasons for their preference. Probing questions were asked when the researcher wanted the students to explain, illustrate, or describe something during the focus groups discussions. Like the individual interviews, the focus group interviews were also conducted in Vietnamese during the school semester from February to May, 2012.

3.6.3. Classroom observation.

Apart from the individual interviews and the focus group interviews, data for the study were also collected from classroom observations. This section provides the rationale for undertaking observations in the EFL classrooms and the type of observation chosen in this study. The section also presents how the observations were undertaken and what was observed and recorded. In addition, the researcher's role in the observations and how rapport between the researcher and the participants and other people in the research sites were established are presented.

3.6.3.1. Reasons for classroom observation.

There were two reasons for conducting classroom observations in this study. First, observations provide knowledge about the context of specific incidents and behaviours (Angrosino, 2007; Merriam, 2009). In this study, observations provided data about the EFL teachers' and their students' assessment activities in the classrooms. Second, data from observations can be used to provide evidence to support or disconfirm the findings from interviews (Merriam, 2009) because interviews "review only what people *perceive* what happens, not necessarily what actually happens" (Bell, 2005, p. 184; emphasis in original). In this study, observations provided evidence of what the teachers and students did around assessment in their classrooms. The data collected from the observations indicated the types of assessment that the EFL teachers implemented in their classrooms, the way they implemented assessment, and the way they negotiated various purposes of assessment in their classrooms.

3.6.3.2. Types of classroom observation.

In order to use observations as a useful data collection method, researchers need to choose a suitable type of observation. Researchers may choose either unstructured or structured observations depending on their purposes of observation. They undertake unstructured observations when they have a clear purpose of the observation but they are not so clear about the detail (Bell, 2005). In other words, researchers undertake unstructured observations to see the "aspects [that] are particularly interesting to their research" (Gibson & Brown, 2009, p. 101). In contrast, structured

observations are chosen when researchers have pre-specified behaviours or practices that they want to know about (Gibson & Brown, 2009). In this study, structured observations of the teachers' classroom assessments were undertaken to discover what assessment practices they used, how they used them, when they assessed their students, and for what purposes they assessed their students.

3.6.3.3. Observing and recording observation.

Observations must be planned, systematic and carefully carried out, recorded, stored, analysed, and interpreted in order to address specific research questions (Angrosino, 2007; Foster, 2006; Merriam, 2009). To meet these criteria, Spradley (1980) suggests that observers should focus their observation on three primary elements: places, actors, and activities. Places refer to space and objects, while actors refer to the people involved in a particular activity. Activities include not only activities, acts, and events but also feelings, perceptions, beliefs, time, and goals. The observation protocols for recording the classroom observations in the study (to be described later in this section) were designed based on these elements.

Although observers need to pay attention to places, people, and activities when they undertake observations, they need to establish the foci for their observations. Several authors have suggested that when undertaking observations, researchers need to focus their attention to and record particular aspects of the observed practices that are useful for answering their research questions (Gibson & Brown, 2009; Sanger, 1996; Spradley, 1980). Following these suggestions, the researcher of this study concentrated his attention on the "*place*, *actors*, and *activities*" (Spradley, 1980, p. 39; emphasis in original) related to classroom assessment, not on all classroom activities. Specifically, he concentrated his observation on the physical features of the classroom, the types of assessment (e.g., tests, examinations, assignments, quizzes), processes of assessment (i.e. how assessments were implemented and how the teacher and his or her students interacted), and timing of assessment (i.e. whether assessment was integrated or separate from instruction).

For each participating EFL teacher, seven classrooms were observed. These classroom observations were undertaken over a period of 15 weeks, from February to May, 2012. A total of 42 observations were undertaken on a pragmatic basis. Three observations per week were feasible for the researcher to manage the collected data. For each teacher, the researcher observed one lesson with a formal test (i.e., a test administered to generate marks for the teacher's grade book) and one lesson in which the teacher returned the students' test papers. The observations of these two lessons provided data about how the teachers implemented formal tests in their classrooms and how they provided their students with feedback. The other five observations were undertaken in five different lessons, each focusing on one language skill (i.e., speaking, listening, reading, writing, and

grammar and vocabulary). These classroom observations provided data about whether or not alternative assessments were implemented and how they were implemented in the classrooms.

Thus, for each teacher, only seven out of 45 classes were observed. This was a small percentage (15.5%) of the classes that each teacher taught in the 15-week semester when the data were collected. It is unknown if the observed classes were representative of the teachers' classroom assessment practices throughout a school year. However the types of assessment practices and the types of lessons (e.g., of each skill type) did reflect the types of assessment practices and types of language lessons that were prevalent in these teachers' classes.

Each classroom observation was recorded on a *Protocol for the Classroom Observations* (Appendix C). The protocol was designed following Spradley's (1980) three-element model. The first page of the protocol was for drawing a map of the classroom. The map showed where objects, the teacher, and students were located in the classroom. The second part of the protocol consisted of multiple pages, each of which had a four-column table. The researcher wrote down the time when an assessment activity started and ended in the first column. In the second and third columns, the researcher wrote detailed and accurate descriptions of the teacher's and the students' words, activities, interactions, and feelings observed in an assessment event. The fourth column comprised descriptions of the resources that the teacher and students used during the assessment. In addition to descriptive notes, Creswell (2012) suggests that observers need to take notes of their thoughts that are related to their hunches, insights, or themes emerging during their observation. The researcher wrote his reflections on the last page of the protocol.

Apart from the observation protocols, an audio recorder was used to record what the teacher and students said and their interactions during assessment-related activities in the classrooms. Audio recording is helpful for analysing observational data because the researcher needs a verbatim record of the participants' words (Spradley, 1980). In this study, the audio recordings of the classroom observations assisted the researcher in getting verbatim records of the teacher's and students' words. However, the quality of the recordings was quite poor because the audio recorder, though placed in the middle of the classroom, recorded too much noise from the ceiling fans, from students inside and outside the classroom, and at times, from machine saws operating just metres away from the classrooms. The poor quality of the recordings made transcribing quite time-consuming.

Video recorders were not used to record the actions in the observed classrooms for three reasons. First, video recorders and video operators may influence the teachers' and students' behaviours and activities when they know that their activities are being recorded. Second, video recorders and video operators may influence observers. Researchers have warned that the presence

of video operators or the researcher's operation of the video recorders may distract him or her from observing what is going on in a setting (Caldwell & Atwal, 2005). Third, researchers have expressed concern about technical limitations of video recorders during data collection, such as the angle of the lens and the places where video recorders are set (Mondada, 2012).

For convenience of analysing and quoting, each observational protocol was assigned a code. TA1Ob.L is used as an example. In this code, T stood for Teacher, A for school A, 1 for the number assigned to one of the three teachers in each school, Ob for Observation, and L for Listening. The letters L, LF, R, S, T, TR, and W stood for Listening, Language Focus, Reading, Speaking, Testing, and Test Return, respectively. The classroom observations were inductively analysed using qualitative content analysis.

3.6.3.4. Observer role and rapport building.

Before conducting the research, the researcher had a limited personal relationship with the principal of the private school (i.e., Principal B). However, the teachers in the schools were unknown to the researcher before the study began. Therefore, there was no conflict of interest with respect to School A, and the researcher had no pre-existing knowledge about the assessment practices in these schools.

An observer may take on different roles depending on the purpose, length, and setting of a study (Teddlie & Tashakkori, 2009). In addition, observers' roles are also influenced by "the degree of their involvement, both *with* people and *in* the activities they observe" (Spradley, 1980, p. 58; emphasis in original). Most authors refer to four roles that range along the participant-observer continuum: complete participant, participant as observer, observer as participant, and complete observer (Angrosino & Mays de Pérez, 2000; Merriam, 2009; Teddlie & Tashakkori, 2009). In this study, the researcher took the role of an "observer as participant" (Angrosino & Mays de Pérez, 2000, p. 113; Merriam, 2009, p. 124) or "participant observer" (Spradley, 1980, p. 53). The observer as participant is involved in "the daily life of an individual, group or community and listening, observing, questioning and understanding (or trying to understand) the life of the individuals concerned" (Bell, 2005, p. 186).

The researcher in this study chose to be a "passive participant" (Spradley, 1980, p. 59). In passive participation, the observer "is present at the scene of action but does not participate or interact with other people to any great extent" (Spradley, 1980, p. 59). More specifically, the observer finds "an 'observation post' from which to observe and record what goes on. If the passive participant occupies any role in the social situation, it will only be that of . . . 'spectator'. . ." (Spradley, 1980, p. 59). During the classroom observations, the researcher was sitting at the back of

the classroom, where he could observe the teacher's and students' activities and hear their words. The researcher did not participate in any activities in the observed classrooms.

In order to get along with the participants and people involved in the research site, researchers need to establish trust and show them respect. Researchers should be helpful (Angrosino, 2007) and also build rapport with the participants and other people in the research site. With trust and rapport, the participants were more open and ready to provide the researcher more information (Bailey, 1996). In order to build rapport with the principals, teachers, and their students, the researcher needs to explain the purpose of his or her presence in their schools and classrooms (Angrosino, 2007). Following these authors' advice, the researcher clearly explained to the gatekeepers and participants the reasons for his presence and activities in their schools and classrooms. Moreover, he behaved in a friendly and respectful way to the school administrators, teachers, and students in the schools.

3.6.4. Documents.

In addition to individual interviews, focus group interviews, and observations, documents were a source of data for this study. Documents refer to "a wide range of written, visual, digital, and physical material" (Merriam, 2009, p. 139). In this study, documents included policy documents which were related to assessment in schools and which were issued by the MOET and the DOET. The English curriculum for Grade 10 to 12 students and the documents related to professional development in assessment for teachers were also used.

Researchers may use relevance or purposive sampling to collect relevant documents for analysis (Krippendorff, 2004). Specifically, they choose texts which are "relevant for their purpose" (White & Marsh, 2006, p. 36). This sampling approach was used to collect the documents to be used as data in this study.

The policy documents concerning assessment are available on the MOET's website (www.moet.gov.vn). The relevant documents were systematically searched using the following procedures.

First, the words *kiểm tra* (assessment, test, or testing), *đánh giá* (evaluation), *thi* (exam, examination, contest, competition) were used as key words to search relevant documents on the MOET's website. The search was limited to those documents which were issued by the MOET since January 1, 2000. The reason for collecting the documents issued since 2000 was that in 2000 the National Assembly issued a resolution to reform school education in Vietnam (National Assembly, 2000).

One hundred and thirty-five documents were found to include these words. Of these 135 documents, some were related to areas which were irrelevant to the present study (e.g., adult education, tertiary education, distance education), so a further selection from these documents took place using two criteria:

1. It was about assessment, testing, tests, evaluation, and/or examinations in classrooms and/or schools.

2. It was relevant to mainstream high school students (Grades 10 to 12).

When the first criterion was applied, those documents which were related to the GEDE and the NUEE were excluded because these national examinations did not belong to school/classroom assessment and the regulations about the examinations were not relevant to school/classroom assessment. The first criterion also excluded those documents which were issued for tertiary, distance, continuing, and adult education. The second criterion excluded the documents related to national examinations for "excellent" students (i.e., national examinations for talented students across the country). Some other documents were also excluded because they were about extracurricular activities (e.g., singing contests), curriculum evaluation, textbook evaluation, school evaluation, online competitions (e.g., in Mathematics and English), and other topics unrelated to assessment. In total, only nine out of the 135 documents satisfied the criteria for collection.

Second, the researcher skimmed the nine documents which satisfied the collection criteria and found that four of the documents had been replaced by four more recently-issued documents. Therefore, the four recently-issued documents were included in the data set for analysis, while the four previously-issued documents were not included in the data set for analysis. This was because the four recently-issued documents made the documents that they replaced officially invalid. This meant that the search for policy documents related to assessment on the MOET's website resulted in five documents directly related to assessment in high schools for analysis. These five documents and a document about assessment which was issued by Binh Dinh DOET and which was provided by one of the principals were put together in a folder called Lot 1. Therefore, there were six policy documents in Lot 1, and these were directly related to assessment.

Following up the principals' suggestions and the references in the documents in Lot 1, the researcher carried out another systematic search of the MOET's website. This time the researcher searched for policy documents about the responsibilities of school staff in each school year since 2000. The word *nhiệm vụ* (responsibility) and *năm học* (school year) were used. The search resulted in 33 documents, and either or both of the following criteria were applied in the selection of relevant documents:

1. The document targeted high schools or high schools belonged to its target. (Some documents targeted not only high school teachers but primary and secondary school teachers as well).

2. The document was about the academic responsibilities of provincial, district, and school administrators and teachers and not about other topics such as national defence education or school inspection.

When the first criterion was applied, nine of the 33 documents were excluded. The second criterion excluded another six documents. As a result, 18 of the policy documents related to the responsibilities of school staff were chosen for analysis and these documents were put together in a folder called Lot 2. Therefore, Lot 2 comprised 18 policy documents directly related to the responsibilities of school staff in each school year and indirectly related to assessment.

When the researcher was gathering other data in the schools in Vietnam, one of the principals provided him with a hard copy of *The professional development document on constructing tests and test banks for administrators and teachers: English language in high schools* (MOET, 2010e). This document was intended to help high school administrators and teachers of English construct multiple-choice items (MOET, 2010e). This document was placed in a folder called Lot 3.

Many of the documents in Lot 1 and 2 also referred to policy documents about professional development (*bồi dưỡng)* for teachers. Most of these policy documents suggested that professional development was an important part of the reform of assessment in schools. Another systematic search of the MOET's website was carried out for policy documents about professional development related to assessment for high school teachers in general and for EFL teachers in high schools in particular. The search resulted in 17 documents, but only 2 documents were chosen after the following criteria were applied:

1. The document was related to high schools.

2. The document was about assessment.

The two policy documents identified in this search were added to Lot 3. However, these were two policy documents rather than professional documents. Specifically, one document, *Circular on issuing the Continuing professional development programme for high school teachers* (MOET, 2011a), was a circular. The other document, *The continuing professional development programme for high schools teachers* (MOET, 2011e), list the modules that high school teachers could choose for their professional learning. These were put in Lot 2. Therefore, Lot 3 consisted of three documents, but only one was a professional development document; the other two documents were policy or official documents related to professional development.

During data collection, the principals provided the researcher with *The English curriculum for high school students* (Binh Dinh DOET, 2009). This document was put in a folder called Lot 4. The principals informed the researcher that this document was a detailed version of the English curriculum for school education issued by the MOET in 2006. The DOET's curriculum and the references in the policy documents in Lots 1, 2, and 3 motivated a search for policy documents about the curriculum for school education. The phrase *chương trình giáo dục phổ thông* (curriculum for school education) was used for the search. The search resulted in three documents related to the curriculum for school education. One document, *Decision on issuing the School education curriculum* (MOET, 2006a), was an official document declaring the issuance of the curriculum for school education. Another document, *The school education curriculum - General issues* (MOET, 2006e), introduced general issues about the new curriculum for school education. The last document was *The school education curriculum for English language* (MOET, 2006f). These three documents were put in Lot 4. Therefore, Lot 4 consisted of four documents related to the curriculum.

In addition to the documents in Lots 1 to 4, two other documents were included in the data set for analysis. These documents were included because they were frequently referred to in the other documents that had been selected. The first document was *Resolution of the National Assembly of the Socialist Republic of Vietnam on reforming the school education curriculum* (National Assembly, 2000). The other document was *Direction for reforming the school education curriculum to execute the Resolution 40/2000/QH10 of the National Assembly* (Prime Minister, 2001). These two documents set out the call for reform in school education and especially in assessment. These two documents were put in a folder called Lot 5.

In total, 33 documents related to regulations about assessment, reform in assessment, professional development in assessment, and curriculum were analysed for this study.

This section has presented the instruments and why and how these instruments were used to collect data to answer the research questions in the study. The next section will present how the collected data were stored and managed for easy retrieval and analysis. The issues of confidentiality of the participants and the schools involved in the study are also addressed in the next section.

3.7. Data Management and Analysis

In order to be secure and useful, the collected data needed to be safely stored and systematically analysed. This section provides a discussion about how the collected data were stored and analysed.

3.7.1. Data management.

The collected data were carefully stored for security and convenience purposes. The interview audio files, interview transcripts, observation audio files, observation protocols, and documents are securely stored and kept confidential. The audio file and transcript of the interview with each principal were stored in one computer folder. Similarly, the audio file of the interview with each teacher, the transcript of the interview, and seven audio files of seven classroom observations of each teacher were stored in one computer folder. The two audio files of the two focus group interviews and the transcripts were stored in one computer folder. The computer folder of the materials collected from a teacher and the folder of the materials collected from his or her students were stored on one master computer folder. In total, there were eight master folders for two principals and six teachers and their students.

The seven classroom observation protocols of each teacher were stored in one paper folder. Therefore, there were six paper folders for six teachers. The policy documents were stored in another paper folder.

All the computer files, folders, and master folders were labelled with pseudonyms and could only be accessed with a password. The paper folders were also labelled with pseudonyms for confidentiality. The computer files and classroom observation protocols that were used for data analysis were the copies of the original files and protocols. The original computer files and classroom observation protocols were stored separately.

3.7.2. Vietnamese-English translation.

The interviews with the principals, EFL teachers, and students were conducted in Vietnamese. In addition, most of the policy documents were written in Vietnamese. Therefore, the transcripts and relevant sections of the policy documents were translated into English for analysis and report. In order to ensure that the English text of Principal A's interview was an accurate translation of the Vietnamese, the researcher sent the transcript of the interview to a colleague and asked her to translate it into English. This colleague had been teaching English to university students, both English majors and none-English majors for 15 years. She had also worked as a freelance translator translating literary works written in English into Vietnamese for several publishers in Vietnam. When compared, the researcher's translation of the transcript of the interview with Principal A and the freelance translator's translation was similar, except that the freelance translator used the word "testing" to mean both testing and assessment. The researcher translated the other interviews and the documents himself.

3.7.3. Data analysis.

Data analysis is "the process of making sense of the data" (Merriam, 2009, p. 175). In this process, the researcher consolidates, reduces, and interprets what participants have said and what the researcher has observed and/or read (Merriam, 2009). Content analysis can be used to analyse interviews, observations, and documents (L. Cohen, Manion, & Morrison, 2007; Merriam, 2009). Therefore, in the present study, content analysis was used to analyse the policy documents, individual interviews, focus group interviews, and classroom observations.

Content analysis is defined as "a systematic method of categorising and analysing the content of texts" (Steenkamp & Northcott, 2007, p. 12). Content analysis is used "to describe the phenomena of interest for a particular purpose" (Downe-Wamboldt, 1992, p. 316), and it helps the researcher gain insights into a particular phenomenon (Steenkamp & Northcott, 2007).

In content analysis, researchers may analyse texts qualitatively or quantitatively, and they may follow an inductive approach or deductive approach (Graneheim & Lundman, 2004; Hsieh & Shannon, 2005). Researchers who use qualitative content analysis and follow the inductive approach read through a text and develop categories and themes that emerge from the text (Elo & Kyngäs, 2008; D. L. Morgan, 1993; White & Marsh, 2006). Their analysis is not "form-oriented" (i.e., counting words, concepts, or themes), but it is "content-oriented" (i.e., focusing on inferences) (Steenkamp & Northcott, 2007, p. 12). When a "content-oriented" approach to content analysis is adopted, "[c]ontent analysis is more than a counting game; it is concerned with meanings, intentions, consequences, and context" (Downe-Wamboldt, 1992, p. 314) and qualitative content analysis is concerned with "latent level analysis" and "the underlying deeper meaning of the data" (Dörnyei, 2007, p. 246).

In content analysis, various units such as words, phrases, sentences, paragraphs, and even whole texts can be units of analysis (Downe-Wamboldt, 1992; Steenkamp & Northcott, 2007). In this thesis, text segments from the documents, interviews, and classroom observations were regarded as the units of analysis.

The researcher followed five steps in analysing the policy documents, interviews, and classroom observations. The steps and procedures used to analyse the policy documents can be seen as illustrations of how the policy documents, interviews, and classroom observations were analysed in this thesis.

In the first step, the researcher familiarised himself with the data by reading the selected documents, one document after another. The first reading of each document served two purposes. The first purpose was to locate, copy, and paste the text segments related to assessment into the second column (*Text Segment*) of a 4-column table (Appendix N). The reference of the document

was typed in the first column (*Reference*) of the table. Locating, copying, and pasting relevant text segments of each document into the table was necessary because not all the text in a document was about assessment. The second purpose was to get a "feeling" or make a tentative conclusion related to assessment. For example, after the first reading of the "Regulation on assessing and ranking secondary and high school students" (MOET, 2011f), the researcher formed a tentative conclusion that policy-makers wanted teachers to assess their students in order to rank them using marks.

In the second step, the researcher read the text segments in the table, one after another, and asked questions, such as "What does this segment of text suggest?" and "What can be inferred from this text segment?". The answers to such questions helped form the categories in the analysis. When the researcher thought of a category name for a text segment, he typed the category name in the 3rd column (*Category*) of the table.

The process of reading, asking questions, finding category names for the text segments and/or matching the text segments with existing category names was repeated for all the text segments, one after another. Sometimes the researcher stopped this process to check and compare the content of the text segments organised in the same category. When a text segment did not match a category, it was moved to another existing category, if that category existed. If no relevant category existed, a new category name was created.

When all the documents had been analysed (that is, all the relevant text segments in the table had been given category names), the researcher started the third step. In this step, the researcher examined the categories to determine whether or not some categories were related to each other. Those categories which were related to each other were grouped together and identified under a common theme in the 4th column (*Theme*) of the table. Again, asking questions around the relationships between the related categories helped group them and assisted in identifying them as themes. General questions such as "How is this category related to other categories?", "Which categories can go together, and why can they go together?", and "what do they suggest?" were asked to group categories into themes.

The researcher started the fourth step when all the categories had been grouped into themes. The researcher examined the relationships between the themes and grouped them into theme clusters.

After the theme clusters had been formed, the researcher took the fifth step: drawing a diagram showing the relationships between the categories, the themes, and the theme clusters. The results of the analysis were now ready for reporting.

The researcher chose to analyse the data using Microsoft Word instead of software such as NVivo. Specifically, the commands in Microsoft Word such as *Find*, *Convert Text to Table*, *Sort*, and *Line Numbers* can be used effectively to analyse the qualitative data.

This section has addressed how the data were stored after being collected and how the data were analysed for reporting. The next section presents the procedures for ethical clearance and recruitment of the participants for the study.

3.8. Ethical Clearance and Participant Recruitment

Bearing in mind the effects that the researcher's data collection and report may have on the schools and participants, the researcher applied for ethical clearance to carry out the research from the Ethical Review Committee of the School of Education, The University of Queensland. When this ethical clearance had been granted, the researcher met with the Head of the DOET of the province where the data were collected. At this meeting, the researcher presented the Head of the DOET with the *Information Letter to the Head of the DOET* (Appendix I) and asked for permission to collect data in two high schools. With the *Permission to Conduct Research from the Head of the DOET* (Appendix M), the researcher met with the principals. Indeed, through the personal relationship between the researcher and the principals, the principals had agreed to allow the researcher to conduct the study in their schools provided that the Head of the DOET granted permission. At the meetings with the principals and the heads of the English Section (i.e., the head of the group of all the EFL teachers in a school) in their schools, the researcher presented the *Participant Information Sheet* (Appendix H) and the *Information Letter to the Principals* (Appendix J) and officially obtained permission to conduct the study in their schools. The principals were also invited to participate in the study. Therefore, they were presented with the *Information Letter to the Principals as Participants* (Appendix K) and the *Consent Form for Interviewees* (Appendix E). The heads of the English Section forwarded copies of the *Participant Information Sheet* (Appendix H) and copies of the *Information Letter to the EFL Teachers* (Appendix L) to their EFL teachers.

With the help of the principals and the heads of the English Section, the researcher had a meeting with the EFL teachers who volunteered to participate in the study. At the meeting, these EFL teachers signed the *Consent Form for Interviewees* (Appendix E) and *Consent Form for Classroom Observations* (Appendix F).

With the teachers' approval, the researcher met their students and told them about the researcher's activities in their classrooms and what they would do if they were chosen to participate in the study. The students were asked to forward copies of the *Participant Information Sheet* (Appendix H) and *Consent Form for Parents* (Appendix G) to their parents for approval and then

give them back to the researcher. All the students in six classes in two schools returned the copies of the *Consent Form for Parents* with either their father's or mother's signature.

This chapter has presented the instruments for collecting the data for the study and how the data were stored and analysed. The chapter has also presented the procedures for seeking ethical clearance and recruiting the participants for the study. The next chapter discusses the results of the analysis of the policy documents and the documents for professional development for EFL teachers in assessment.

Chapter 4. Assessment as Constructed in Policy Documents

4.1. Introduction

This chapter discusses the results of the analysis of 32 documents. These documents included 29 policy documents directly or indirectly related to assessment, two policy documents related to reform in assessment, and one professional development document that the MOET designed to provide techniques for constructing multiple-choice items and tests for high school administrators and EFL teachers. The procedures for collecting and analysing these documents were presented in Chapter 3.

The results of the analysis of the documents are summarised in Figure 1. There were three theme clusters. The first cluster comprised two themes related to reform in assessment. The first theme indicated that there was a need for reform in assessment in school education in Vietnam. The second theme showed what policy-makers expected from reform in assessment. The second cluster was made up of three themes, indicating that the MOET wanted to preserve a culture of testing. The first theme revealed that policy-makers expected that students' learning outcomes were assessed against the MOET's standards in equitable conditions. The second theme showed the purposes that the MOET expected of assessment. The third theme in the second cluster was about professional development for EFL teachers in assessment. The third cluster included two themes indicating that the MOET promoted a culture of assessment. The first theme was related to the purposes of assessment that the MOET promoted, and the second theme was about the types of assessment practices that the MOET expected school teachers to implement.

4.2. Reform in Assessment

The analysis of the policy documents related to assessment in schools and *The professional development document on constructing tests and test banks for administrators and teachers: English language in high schools* (MOET, 2010e) indicated that policy-makers realised the need for reforming assessment in school education in Vietnam. They also expected that the reform in assessment would bring about positive changes in the methods of teaching and learning, in improving the quality of teaching and learning, and in achieving objectivity, accuracy, and fairness in assessment.

Figure 1. Results of the Analysis of the Policy Documents

Assessment as Constructed in Policy Documents		
Reform in Assessment	**Preserving a Culture of Testing**	**Promoting a Culture of Assessment**

Clusters

Need for the reform in assessment - Integral part of reform in school education - Central duty of school staff - Teachers' inadequate competence in assessment **Policy-makers' expectations of positive changes from reform in assessment** - Promoting changes in methods of teaching and learning - Improving the quality of teaching and learning - Ensuring accuracy, objectivity, and fairness in assessment	**Measuring learning outcomes in equitable testing conditions** - Standards-based assessment - Product-oriented assessment - Equitable testing conditions **Generating marks for ranking, promoting, and retaining students** - Summarising learning outcomes - Ranking students - Promoting or retaining students **Providing teachers with techniques for constructing tests**	**Assessment for the improvement of teaching and learning** **Call for use of multiple types of assessment** - Encouraging teachers to use multiple types of assessment - Emphasising self-assessment and peer-assessment

Themes and Categories

Figure 1. Categories, themes, and theme clusters emerging from the analysis of the MOET's and the DOET's policy documents and professional development document related to assessment in schools

4.2.1. The need for reform in assessment.

There were three manifestations of the need for reform in assessment. First, reform in assessment was seen as an integral part of the reform in school education. In the *Resolution of the National Assembly of the Socialist Republic of Vietnam on reforming the school education curriculum*, which approved the government's policy on reforming the school education curriculum, the National Assembly mandated, "The reform in the curriculum, textbooks, and methods of teaching and learning has to be carried out in parallel with . . . the reform in assessment" (National Assembly, 2000, p. 1). The need for reform in assessment was also emphasised in the Prime Minister's *Direction for reforming the school education curriculum to execute the Resolution 40/2000/QH10 of the National Assembly* (Prime Minister, 2001). The Prime Minister's direction emphasised, "Synchronise the reform of the curriculum, textbooks, and methods of teaching and learning with the fundamental reform in the methods of assessment . . ." (pp. 1-2). The reform in assessment was "a requirement that needs to be fulfilled when the reform in the methods of teaching and learning and the reform in education [in general] are undertaken" (MOET, 2010e, p. 4). These regulations indicated that reform in assessment was seen as an inseparable part of reform in school education.

Second, because reform in assessment is seen as "an important long-term task" (MOET, 2010e, p. 7) and "both an urgent and long-term task" (MOET, 2010e, p. 12), it has been regulated as one of the central responsibilities of school staff for over a decade now. The requirement for reform in assessment has been manifested in the MOET's regulations on school staff responsibilities in each school year since 2002. In 2002, the MOET required school staff to "continue changing examinations and methods of assessing student learning outcomes" (MOET, 2002, p. 1). A similar requirement was repeated in 2003: "[R]eform assessment of student learning outcomes" (MOET, 2003a, p. 2). In 2004, the MOET had the same requirement and added some specific tasks that school staff needed to do.

> Continue changing and step by step turn the practice of assessment, recruitment, and evaluation of the teaching and learning process into a stable discipline at all educational levels. . . . firmly remedy the status of assessing students only to have the mandatory number of marks [for each student]. . . (MOET, 2004, p. 4).

The direction "reform the practice of examination, assessment, evaluation" was repeated in three successive years (MOET, 2005a, p. 5; 2006b, p. 2; 2007, p. 6). Although the wording was slightly different, the same direction, "reform assessment", has been given to school administrators and teachers as one of their central responsibilities for many years (MOET, 2008a, p. 12; 2009c, p. 10; 2010a, p. 1; 2011c, p. 2; 2011d, p. 1; 2012a; 2013, p. 5; 2014, p. 5).

Third, the need for reform in assessment was the result of the MOET's perception that school teachers did not have appropriate competence in assessment. According to the MOET, although assessment was an important part of instruction, most school teachers were not prepared for it.

> . . . in reality, most teachers were not provided with techniques for assessment at teacher training colleges . . . . Many teachers have met difficulty in designing multiple-choice tests, leading to the fact that the quality of multiple-choice tests is not high, not appropriate to the tested content areas and the particularities of [their] subject, multiple-choice tests have been overused in many cases (MOET, 2010e, p. 8).

These judgements about teachers' training and competence in assessment indicated that the MOET distrusted teachers' skills in constructing and using multiple-choice tests. They saw teachers' lack of skills in assessment as endemic, arising out of poor training in assessment. There was no mention of continuing professional development for teachers to enhance their assessment literacy in the policy documents. However, the principals' beliefs about the teachers' lack of training and competence in assessment might suggest that teachers could benefit from professional development to improve their knowledge and skills in assessment.

The discussion in this sub-section indicated that policy-makers believed that there was a strong need for reform in assessment in school education in Vietnam. Policy-makers expected that reform in assessment would bring about significant changes to school education in the country.

4.2.2. Policy-makers' expectations of positive changes from reform in assessment.

The MOET had two expectations from reform in assessment. First, they expected that reform would promote changes in methods of teaching and learning. This expectation was repeated in various documents. The MOET directed school administrators to "[s]teer reform in assessment in order to promote reform in methods of teaching and learning" (MOET, 2009c, p. 10). In addition, the MOET directed school staff to hold "conferences on reforming assessment in order to promote reform in methods of teaching and learning" (MOET, 2010a, p. 8). In 2011, the MOET required school administrators to "[f]ocus on directing the promotion of the effectiveness of reforming assessment in order to promote reform in methods of teaching and learning" (MOET, 2011d, p. 1). These regulations indicated that the MOET perceived reform in assessment to be a driving force of reform in methods of teaching and learning.

Second, the MOET expected that reform in assessment would improve the quality of teaching and learning. This expectation was manifest in the professional development document. ". . . reform in methods of teaching and learning and reform in assessment are the key solutions to increasing the quality of teaching and learning in particular and the quality of comprehensive education in general" (MOET, 2010e, pp. 11-12). The same expectation was found in the MOET's

directions for school staff. The MOET directed school staff to "make basic changes in assessment to . . . increase the quality of education" (MOET, 2010a, p. 1). These statements indicated that policy-makers had high expectations of positive changes in the quality of teaching and learning resulting from reform in assessment.

Third, the MOET expected that reform in assessment would help achieve objectivity, accuracy, and fairness in assessment. They regarded the achievement of these qualities as "an important requirement in reform in assessment" (MOET, 2008b, p. 5). The MOET stipulated, "Teachers must assess students' performance accurately with an objective, fair, and transparent attitude . . ." (MOET, 2008b, p. 5). The same requirement was also found in the professional development document, "When reform in assessment is implemented, it is necessary to ensure the requirement of objectivity, accuracy, [and] fairness" (MOET, 2010e, p. 6). This statement suggested that ensuring these qualities was an important part of reform in assessment. Specifically, the MOET believed that achieving these qualities "helps teachers and managerial bodies determine the effectiveness of teaching and learning [and] create the basis for teachers to change methods of teaching and learning and for managers at different levels to put forward appropriate managerial solutions" (MOET, 2010e, p. 6). The MOET also believed that there was a link between the achievement of objectivity, accuracy, and fairness in assessment and the improvement in the quality of teaching and learning. "It is necessary to ensure objectivity, accuracy, and fairness in order to raise the quality of teaching and learning" (MOET, 2010e, p. 8). The expectations related to objectivity, accuracy, and fairness indicated that the MOET valued these qualities in assessment.

The discussion in this sub-section indicated that policy-makers expected that positive changes would result from reform in assessment. They expected that reform in assessment would promote change in methods of teaching and learning and improve the quality of teaching and learning. They regarded achieving accuracy in test results and objectivity and fairness in testing as a target of reform in assessment. However, the emphasis on accuracy, objectivity, and fairness suggested that policy-makers were concerned very much about preserving a culture of testing.

4.3. Preserving a Culture of Testing

Elements of a culture of testing were manifest in the regulations about how tests were administered and how test results were used. In addition, a culture of testing was supported in the recommended professional development for EFL teachers, which focused on providing EFL teachers with techniques for designing multiple-choice items and tests.

4.3.1. Measuring learning outcomes in equitable testing conditions.

Assessment mandated by the MOET focused on measuring student learning outcomes. First, the documents frequently referred to assessment as standards-based testing. The MOET required that tests were designed according to the MOET's standards of knowledge and skills of the subjects. The MOET required school staff to "assess [students] according to the standards of knowledge and skills [defined] in the school education curriculum" (MOET, 2011c, p. 2).

Second, assessment was expected to be product-oriented. The MOET directed that schools "intensify external assessment" (MOET, 2011c, p. 2). That is, "schools can invite another school [or] a professional organisation to assess their students' learning outcomes" (MOET, 2010e, p. 5). In reality, high schools in Vietnam designed tests for their own students. External examinations included only the GEDE and the NUEE because these were designed by the MOET and administered by the DOETs and a number of universities appointed by the MOET. Therefore, the suggestion about external assessment carried out by "another school" or "a professional organisation" suggested that the MOET emphasised assessment of students' learning outcomes rather than their learning processes. Thus, assessment of learning rather than assessment for learning was valued.

The regulation on the timing of 45-minute tests and end-of-term tests (Binh Dinh DOET, 2010) clearly indicated that assessment was product-oriented. According to this regulation, teachers had to administer tests in the time frame set by the DOET. A prescribed time frame for testing revealed that assessment was separated from instruction. This meant that assessment was not intended to assess students' learning processes.

Third, the DOET's mandate that students in the same grade took the same end-of-term test at the same time indicated that they expected equitable conditions in testing.

> - When administering end-of-semester tests, schools have to list students in a, b, c order in each grade block [all the students in the same grade in a school] when they take common tests in high schools (Grades 10, 11, 12) and no more than 45 students are to be seated in each room.
> - When test papers are marked, they have to be anonymous and teachers gather to mark the test papers . . . . (Binh Dinh DOET, 2010, p. 2)

The provision of equitable conditions to test-takers showed that assessment was high-stakes (Estaji, 2011). In addition, the regulation about anonymous marking suggested that policy-makers distrusted teachers' professional integrity. The DOET's regulations about equitable conditions in testing and anonymous marking indicated that accuracy of marks as measures of learning outcomes and teachers' objectivity and fairness were valued.

The requirement of equitable conditions in testing were so highly valued that the MOET mandated that a student took a substitute test with the same format, similar difficulty, and the same time allocation if she or he missed a test.

> Those students who have not taken enough tests as mandated . . . have to take substitute tests. The substitute tests must have the same form, the same level of knowledge and skills, and the same time allocation as the one they have missed. Those students who do not take substitute tests are to be given a mark of zero . . . . (MOET, 2011f, p. 7)

The above regulation indicated that measuring student learning outcomes in equal testing conditions was of great concern for policy-makers. Not only did the regulations on assessment administration indicate that assessment focused on measuring learning outcomes, the regulations on the use of assessment data showed that this was the main purpose that the MOET expected teachers to fulfil.

4.3.2. Generating marks for ranking, promoting, and retaining students.

Tests were mandated to be delivered at the end of learning units in order to generate marks for summarising student learning outcomes. This was described in the *Guidelines on designing tests* (MOET, 2010b). "Tests are tools used to assess student learning outcomes after they have covered a topic, a chapter, a semester, a year, or an educational level [e.g., primary, secondary, and high school education]" (MOET, 2010b, p. 3). Although the use of various methods of assessment was encouraged, in English instruction, for example, reference was made only to tests. "Methods of assessment need to be diverse, including oral tests, 15-minute tests, one-period tests, end-of-term tests, and end-of-year tests" (MOET, 2006f, p. 18). The purpose of summarising students' learning outcomes could also be seen in the references to the timing of tests (i.e., when tests were administered). The DOET of the province where the data for this study were collected mandated that 45-minute tests, end-of-term tests, and end-of-year tests be implemented when two or three units of work had been covered. Specifically, for English language, the tests were set for the 19th, 37th, 52nd, 70th, 88th, and 104th period for Grades 10, 11, and 12 (Binh Dinh DOET, 2009). This regulation had the potential to affect EFL teachers' autonomy in deciding when to implement tests, thus restricting their use of assessment for supporting teaching and providing feedback for students to improve their learning. The purpose of summarising student learning outcomes was also reflected in teachers' responsibilities in assessment. The *Regulations on assessing and ranking secondary and high school students* (MOET, 2011f) specified teachers' responsibilities as follows.

> 1. Implement the mandatory number of tests, mark test papers, assign marks . . ., write comments on the test papers, input marks . . . into the class register; as for oral tests, the teacher has to give comments on the students' answers in front of the class, if the teacher decides to give a mark . . . in the class register, s/he has to write down the mark immediately.

2. Calculate the average mark for the subject . . . for each semester, for the whole school year and write down [the average marks] in the class register and in the students' records. (p. 10)

Although providing feedback (i.e., comments) was mentioned in teachers' responsibilities related to assessment, these extracts suggested that most of teachers' tasks in assessment were related to tests and marks. Moreover, administering tests and generating marks for summarising student learning outcomes were an important role of assessment that the MOET expected teachers to fulfil.

Generating marks was important because they were also used to rank students. The *Regulations on assessing and ranking secondary and high school students* (MOET, 2011f) stipulated,

1. The basis for evaluating and ranking learning competence:
a. How much of the curriculum for the subjects and educational activities in the Educational plan for secondary and high schools that the student has completed.
b. The test results.
2. Learning competences are classified into 5 rankings: Outstanding, good, pass, poor, and very poor. (p. 3)

In effect, how much of the curriculum a student completed was just one factor in judging his or her learning outcomes; the ranking s/he was given was considered to be the decisive factor.

Marks were also used for making decisions about letting students move up to the next grade or repeat their grade. A student's ranking became the basis for a decision about whether the student was allowed to move up to the next grade or to repeat the grade (MOET, 2011f). The MOET required that assessment needed to be undertaken according to regulations because assessment results were used to ensure that only those students who achieved the outstanding, good, and pass rankings were allowed to move up. "Assess and rank student learning outcomes according to regulations. . . [and] ensure that students [ranked as outstanding, good, and pass] are allowed to move up to the next grade . . ." (MOET, 2007, p. 3). The regulations on how test results were used indicated that marks had high-stakes consequences for students.

The MOET and the DOET mandated that assessment be administered to generate marks to rank students and decide whether students were qualified to move up to the next grade or they had to repeat the grade. Therefore, teachers' tasks in assessment were mainly related to tests and marks. Test results were high-stakes to students. The focus on measuring student learning outcomes appeared to be supported by the emphases found in the document for EFL teachers' professional development in assessment.

4.3.3. Providing teachers with techniques for constructing tests.

The results of the analysis of *The professional development document on constructing tests and test banks for administrators and teachers: English language in high schools* (MOET, 2010e)

and other policy documents related to professional development indicated that the MOET focused on providing teachers with techniques for designing multiple-choice items and tests.

In *The continuing professional development programme for high schools teachers* (MOET, 2011e), the two modules designed to "[i]ncrease the competence in assessing student learning outcomes" (p. 8) focused on techniques for constructing multiple-choice items and tests. These two modules were part of a 41-module professional development programme that the MOET required high school teachers to study. (The other 39 modules were not related to assessment.) In the first module, which was designed to help teachers "[d]ifferentiate and implement the methods of assessing student learning outcomes" (MOET, 2011e, p. 8), the following content areas were listed without any elaboration.

1. The role of assessment
2. Methods of assessing student learning outcomes
3. Using assessment methods to assess student learning outcomes. (p. 8)

These content areas suggested that the MOET wanted teachers to assess students' learning outcomes rather than using assessment to support teachers and to provide feedback for students to improve their learning.

The other module was intended to help high school teachers "[u]se assessment techniques in teaching" (MOET, 2011e, p. 8). This module, titled "Techniques for assessment in teaching", included the following topics:

1. Techniques for designing tests: determining the objective of the test; form of the test [whether the test includes open-ended items or multiple-choice items or both open-ended and multiple-choice items]; designing the test matrix; writing the test and designing the rubric.
2. Techniques for analysing test results in order to improve the effectiveness of teaching and learning. (p. 8)

These content areas focused entirely on techniques related to tests and testing rather than assessment in general. Although the module was also intended to provide teachers with "[t]echniques for analysing test results in order to improve the effectiveness of teaching and learning" (p. 8), there was no mention in the module of the knowledge and skills needed to use the results of their analysis for improving teaching and learning.

The regulation that teachers developed techniques to design tests according to standards was so important that the MOET repeated it in three policy documents issued in three successive years. "Intensify the professional development for teachers in techniques and skills in designing tests, designing rubrics, and assessing students using open-ended and multiple-choice test items according to standards of knowledge and skills in the school education curriculum . . ." (MOET, 2010a, p. 7;

2011d, p. 5; 2012b, p. 5). This regulation revealed that the MOET valued teachers' capacity to construct tests and expected teachers to assess their students against the MOET's standards.

In line with the above regulation, *The professional development document on constructing tests and test banks for administrators and teachers: English language in high schools* (MOET, 2010e) suggested techniques that EFL teachers should use to design test items. Statements such as the following were pervasive in the document. The first statement referred to the construction of an item to assess vocabulary.

> The following steps should be taken in writing multiple choice completion items:
> (1) Select the words to be tested.
> (2) Get the right kind of sentence to put each word in (the sentence creating the context is called *stem*).
> (3) Choose several wrong words to put the right word with (these wrong words are called distractors). Three distractors plus the right word are enough for a written item.
> (4) Finally, prepare clear and simple instructions. And if this kind of test question is new to your students, it would be recommendable to prepare one or two examples. (MOET, 2010e, p. 14, English in original, emphasis in original)

Providing EFL teachers with these very specific guidelines on how to design multiple-choice items to assess vocabulary, the MOET assumed that teachers lacked skills in designing multiple-choice items. These guidelines also suggested that the MOET was concerned about technical aspects of assessment.

The next extract describes how teachers should create test items to assess grammar knowledge.

> *1.1. Hoàn thành câu nhiều lựa chọn (Multiple-choice completion)*
> The test type presented in this part includes an incomplete sentence stem followed by four multiple-choice options for completing the sentence. Here is an easy sample item:
> *E.g.: She is ------------- her breakfast.*
>   *A. eating   B. ate   C. eats   D. eaten*
> While multiple-choice completion is an efficient way to test grammar, teachers need to be cautioned about the temptation to use this bid [sic] of item for all of their testing needs. Many people are very excited about objective tests, feeling that multiple choice objective exams in particular should be used to test everything.
> However, any given test is a kind of tool; it may be very useful for some jabs [sic] but not for others. For example, while multiple-choice tests can be used successfully in testing grammar, they don't seem to work as well in testing conversational ability.
> Preparing multiple-choice completion grammar items follows about the same procedure as that described in the previous part for writing multiple-choice completion vocabulary items: (1) Choose the grammar points that you need to test; (2) prepare the right kind of sentence context (or stem) for the grammar structure; (3) select three logical distractors; and (4) prepare clear, simple instruction [sic]. (MOET, 2010e, p. 22, English in original, italics in original)

This extract suggested when teachers should use multiple-choice items and when they should not. Further, the extract reminded teachers not to overuse multiple-choice tests. However, the main role

of the extract was to provide teachers with specific guidelines on how to design multiple-choice items to assess grammar. The extract indicated that the MOET was concerned about the quality of test items, especially multiple-choice items.

In addition to techniques for designing vocabulary and grammar test items, techniques for designing reading and writing test items and techniques for designing test matrices were provided in the same professional development document.

The discussion in this section has indicated that policy-makers supported a culture of testing. Specifically, they expected teachers to assessment students' learning outcomes according to the MOET's standards of knowledge and skills. They valued objectivity, accuracy, and fairness because assessment was expected to be administered to generate marks. Marks were high-stakes for students because they were used to rank students and make decisions about students' promotion and retention. In addition, a culture testing was valued in that teachers were provided with guidelines on how to design multiple-choice items and tests. The MOET valued the quality of test items in assessing students' listening, reading, writing, and grammar in English. The policy documents and professional development documents did not provide EFL teachers with any advice on using assessment for formative purposes. Nevertheless, the documents did provide EFL teachers with advice on designing test items and tests. The lack of guidelines about how to administer assessment for teaching and learning improvement may have led the EFL teachers to believe that policy-makers prioritised testing students in order to generate marks rather than assessing them for teaching and learning improvement. In addition, this lack of advice may explain why the EFL teachers in this study relied too much on the textbooks in their classroom assessment.

Although the professional development documents for EFL teachers focused on providing them with skills in designing test items and tests, there was evidence that the MOET wanted to promote a culture of assessment.

4.4. Promoting a Culture of Assessment

The MOET wanted to promote a culture of assessment in schools. They expected teachers to use assessment for the purpose of modifying and improving teaching and learning. For this purpose, they expected teachers to implement multiple types of assessment and embed assessment into instruction. In addition, policy-makers expected teachers to encourage students to take active roles in assessment.

4.4.1. Assessment for the improvement of teaching and learning.

The MOET expected teachers to use test results to modify and improve teaching and learning. "Appreciate the analysis of test results, according to which teachers modify their teaching, help students improve their strengths and remedy weaknesses in learning. . ." (MOET, 2009b, p. 2). A similar purpose also can be seen in the following extract.

> Assessing student learning outcomes is a very important activity in the educational process. Assessing student learning outcomes is the process of collecting and processing the information about students' ability to fulfil learning objectives in order to support teachers' pedagogical modification and solutions of managers at different levels and for students themselves so that students can achieve better learning outcomes. (MOET, 2010d, p. 1)

This extract indicated that the MOET expected teachers, managers, and students to use assessment data to improve student learning through changes in teachers', managers', and students' activities. The MOET suggested that teachers should use assessment data and feedback to help students improve their learning. "In instruction and assessment, teachers have to know how to 'exploit errors' to help students see their errors in order to help students with the methods of learning and methods of thinking" (MOET, 2010e, p. 5). However, the above extract revealed that the MOET expected teachers to assess students' learning outcomes rather than their learning processes. When tests were separated from instruction, the purpose of supporting learning through feedback was not feasible. Therefore, the purpose of improving teaching and learning were subordinated to the purpose of generating marks.

4.4.2. Call for use of multiple types of assessment.

Another feature of a culture of assessment was manifest in the the MOET's call for the use of multiple types of assessment.

> Require teachers to reform methods of assessing students. It is necessary to use various assessment types such as research projects; to assess products of students' learning activities (folders of tests that they have done best; folders of pictures that students have collected; folders of essays, poems, articles that students have collected according to topics; students' learning journals . . .); to assess students through performance (playing musical instruments, using machines . . .); to assess students through oral presentations; to assess students through teamwork; to assess students through groups' activities . . . . (MOET, 2010e, p. 11)

The MOET regarded using various methods of assessment as part of reform in assessment. This excerpt indicated that tests were not considered to be the only method of assessment. Instead, teachers were encouraged to use methods of assessment that were typical of a culture of assessment.

The MOET expected teachers to use various types of assessment, depending on the content areas that they wanted to assess and the subject they taught, so that they could gather accurate information about student learning outcomes.

Each method [of assessment] has its strengths and weaknesses, so it is necessary to combine various methods in a way that is appropriate for the tested content areas and the peculiarity of the subject in order to improve the effectiveness [of assessment], creating conditions for accurately assessing student learning outcomes. (MOET, 2010b, p. 1).

The extract indicated that the MOET encouraged teachers to use various methods. However, it suggested that policy-makers expected teachers to pay attention to the product of learning rather than to both process and product of learning. Moreover, the extract suggested that the MOET expected teachers to use various methods of assessment to ensure that assessment results accurately reflected students' learning outcomes rather than to support teaching and to provide feedback for students to learn.

In the call for use of multiple methods of assessment, the MOET emphasised self-assessment and peer-assessment. "Reform in assessment is effective only when teachers' assessment is combined with students' self-assessment" (MOET, 2010e, p. 5). They expected teachers to promote peer-assessment and self-assessment in the classroom: "In assessing student learning outcomes, . . . guide students in assessing each other and assessing their own competence" (MOET, 2011d, p. 4). However, self-assessment was understood as checking a test paper which had been marked by the teacher in order to see whether or not the teacher's marking was accurate. "After each test, teachers need to return the test papers back to students and guide them in checking their own test results, assigning a mark for their own test paper, and commenting on the accuracy of their teacher's marking" (MOET, 2010e, p. 5). This extract indicated that the MOET were concerned about the accuracy of marks rather than how teachers used self-assessment to help students improve their learning.

The analysis in this section indicated that policy-makers expected teachers to implement assessment for the purpose of improving teaching and learning. In order to fulfil this purpose, they expected teachers to use various methods of assessment, especially self-assessment and peer-assessment and to provide students with feedback. These expectations indicated that policy-makers promoted a culture of assessment in schools in Vietnam. However, the MOET was more concerned about the accuracy in assessing students' learning outcomes. They expected teachers to use various methods of assessment. However, their expectation from teachers' use of multiple methods of assessment was to achieve accuracy in measuring students' learning outcomes rather than to support and improve teaching and learning.

4.5. Discussion

The inclusion of the policy documents and professional development documents as a source of data was intended to shed light on policy-makers' expectations concerning assessment at the

school level in Vietnam. The analysis of the relevant documents indicated that policy-makers in Vietnam still wanted to preserve a culture of testing, while they also promoted a culture of assessment. They perceived that assessment in school education needed to be changed and regarded reform in assessment as an integral part of reform in school education. They perceived that reform in assessment was a driving force for reform in methods of teaching and learning and that it would improve the quality of teaching and learning. In addition, they believed that reform in assessment was necessary for the achievement of accuracy, objectivity, and fairness in assessment. Similar to policy-makers in other countries in Asia, such as China (Berry, 2011b), Hong Kong (Berry, 2011a, 2011b; Lee & Coniam, 2013), Taiwan (Berry, 2011b), Malaysia (Ong, 2010), and Singapore (Tan, 2011), policy-makers in Vietnam promoted a shift from assessment of learning to assessment for learning.

Although policy-makers in Vietnam believed that reform in assessment would improve teaching and learning, they still paid much attention to elements of a culture of testing. They focused on assessment of students' learning outcomes rather than both learning outcomes and learning processes for improving learning as suggested by researchers (e.g., Brookhart, 2009). Policy-makers expected teachers to administer tests to generate marks for ranking, promoting, and retaining students. These uses of assessment results indicated that assessment was high-stakes to students. The high stakes or consequences of assessment results were felt by the EFL teachers (see Chapter 6) and the students (see Chapter 7) in this study, and both the teachers and students were more concerned about marks than the formative functions of assessment.

Because assessment results were high-stakes, ensuring accuracy of marks as measures of learning outcomes and achieving objectivity and fairness in assessment were a great concern of policy-makers. This concern was handed down to the principals (see Chapter 5) and the EFL teachers (see Chapter 6), who were very much concerned about tests and test administration. To ensure accuracy, objectivity, and fairness in assessment, the principals restricted their teachers' autonomy in administering assessment. However, these qualities were not achieved because several teachers modified their tests in order to raise their own students' marks (see Chapter 6) and some students cheated to get good marks (see Chapter 7 and 8).

The focus of professional development on providing teachers with techniques for constructing test items and tests, especially multiple-choice items, indicated that policy-makers paid more attention to tests than other methods of assessment, even though they encouraged school teachers to use various methods of assessment, especially self-assessment and peer-assessment. The focus of professional development for teachers suggested that a culture of testing was valued, and

this focus seemed to be related to the teachers' lack of knowledge and skills in classroom assessment (see Chapter 5, 6, and 8).

This chapter also indicated that policy-makers in Vietnam expected teachers to implement alternative assessments, especially self-assessment and peer-assessment, for the purpose of modifying teaching and supporting students' learning. This expectation should be supported by providing teachers with knowledge and skills in implementing alternative assessments and using classroom assessment data to support teaching and to help students improve their learning. However, the MOET's professional development was inconsistent with their expectation. The content of the professional development documents indicated that school teachers were not provided with guidelines about how to use classroom for formative purposes, and the teachers in this study relied on the exercises and tasks in the textbooks (see Chapter 8).

This chapter indicated that policy-makers in Vietnam wanted to preserve a culture of testing and, at the same time, they promoted a culture of assessment. However, their expectations and support for teachers showed that elements of a culture of testing received more attention. In such a context, school principals, teachers, and students may hold different perceptions of assessment in their schools, and assessment practices may reflect prioritised assessment (i.e., assessment of learning or assessment for learning). In order to know the context of assessment in these schools, it is necessary to know the principals', teachers', and students' perceptions of assessment in their school. Therefore, these stakeholders were interviewed and the interviews were analysed and reported. The next chapter will address the principals' perceptions of assessment in their contexts.

Chapter 5. The Context of Assessment as Perceived by the Principals

5.1. Introduction

This chapter discusses the themes and theme clusters emerging from the analyses of the interviews with Principal A (PA) in a public high school (School A) and Principal B (PB) in a private high school (School B) in Vietnam. The categories, themes, and theme clusters emerging from the analysis of the interviews with the principals are set out in Figure 2 which shows that there were three theme clusters. The first cluster showed that the principals perceived that assessment had multiple purposes. The four themes in this cluster revealed the purposes of assessment perceived by the principals. The second cluster was about the principals' perceptions of technical and ethical issues in assessment. The first theme in this cluster indicated that the principals assumed tests to be a technology for measuring students' knowledge. Therefore, they were concerned about tests and test administration. The third cluster was related to the principals' expectation of the EFL teachers' compliance with the regulations and procedures for assessment. The first theme in this cluster showed that the principals expected their teachers to comply with the regulations and procedures for assessment. The second theme revealed the principals' distrust of the teachers' professional competence and commitment as well as their fairness in assessment. This chapter discusses these themes and theme clusters.

5.2. Perceived Value of Assessment and Competing Purposes

The principals' accounts indicated that they valued assessment in teaching and learning. They perceived that classroom and school assessment should help prepare the students for the national examinations. They also believed that assessment, specifically marks, should be used for improving teaching and learning, for managerial purposes, and for accountability purposes. The following sub-sections elaborate the principals' perceptions of the value of assessment and its purposes.

Figure 2. Results of the Analysis of the Interviews with the Principals

The Context of Assessment as Perceived by the Principals

Perceived Value and Purposes	Concerns about Technical and Ethical Issues	Expectations of Compliance and Distrust	Clusters

Valuing assessment for learning and teaching

Preparation for the national examinations
- Passing the national examinations as evidence and target of learning
- Preparing for the national examinations
- Concerns about test-taking skills

Professional purposes
-Informing the teachers for teaching modification and improvement
- Informing the students for learning modification and improvement

Managerial purposes
- Assessment for quality control
- Assessment for school administrators' interventions
- Reporting to parents for their involvement

Accountability purposes
- Controlling learning and making the students accountable
- Controlling teaching and making the teachers accountable

Concerns about technical issues
- Tests as a technology for measuring knowledge
- Concerns about the tests and test administration

Concerns about ethical issues
- Ethical issues associated with the teachers' test construction
- Concerns about objectivity and fairness

Principals' expectations of the teachers' compliance
- Valuing the teachers' compliance
- Compliance with the MOET's and the DOET's regulations
- Compliance with school regulations
- Equating the teachers' knowledge and skills in assessment with their compliance
- Counter-effect of the teachers' compliance

Principals' distrust of the teachers
- Distrust of the teachers' professional competence
- Distrust of the teachers' professional commitment
- Distrust of the teachers' fairness

Themes and Categories

Figure 2. Categories, themes, and theme clusters emerging from the analysis of the interviews with the principals in two high schools

5.2.1. Valuing assessment for teaching and learning.

The principals emphasised the value of assessment in teaching and learning. Principal B stated, "Assessment is an important part of teaching and learning." Similarly, Principal A asserted, "Assessment is indispensable in teaching and learning." To justify the importance of assessment in learning, Principal B said, "The students base their direction in learning on teaching methods and assessment methods. I remember that an educator said that in order to know about a nation's educational system, look at its assessment methods." This statement suggested that Principal B believed that assessment was important to student learning because it influenced how the students learned. He also believed that assessment methods used in a country represented its education system. The principals believed that assessment had multiple purposes. The next sub-section elaborates these purposes.

5.2.2. Competing purposes of assessment.

Researchers have identified multiple purposes of assessment (G. T. L. Brown, 2004; Stobart, 2008; Tanner & Jones, 2003). Many researchers have categorised the purposes of assessment as *formative* and *summative* purposes (Black & Wiliam, 2008; Brookhart, 2001; Harlen & James, 1997) . These two groups of purposes are also called assessment *for* learning and assessment *of* learning, respectively. Formative purposes refer to the purposes of supporting and improving learning and teaching. Other purposes of assessment such as selection, certification, classification, and accountability are subsumed in summative purposes (Black & Wiliam, 2008; Brookhart, 2001; Harlen, 2007; Harlen & James, 1997).

Researchers do not agree on the categories used to organise the purposes of assessment. Buhagiar (2007) refers to selection and certification as managerial purposes, while Tanner and Jones (2003) categorise selection and certification as managerial and communicative purposes, respectively. Buhagiar (2007) refers to accountability as an independent purpose related to using assessment "to evaluate teachers, schools, or age groups at national level" (p. 45), while Tanner and Jones (2003) consider accountability to be one of the managerial purposes. Buhagiar (2007) uses professional purposes to refer to the formative use of assessment for enhancing learning and teaching. In this thesis, "managerial purposes" refer to the school administrators' use of assessment results for quality control and for making decisions about investment in facilities and collaboration with parents in helping the students learn. Following Buhagiar (2007), "professional purposes" refer to the teachers' and students' use of assessment for supporting and improving learning and teaching. In this thesis, "accountability purposes" are related to the use of assessment to control the teachers and students and to make them accountable for teaching and learning.

The principals' accounts referred to four purposes which competed with each other. First, they perceived that one of the purposes of the 45-minute and end-of-term tests was to prepare the students for the national examinations. Second, assessment helped improve teaching and learning (i.e., professional purposes). The professional purposes competed with the managerial purposes and accountability purposes. While the professional purposes require students' active participation, timely and constructive feedback from teachers and peers, and alternative assessments (Black & Wiliam, 1998a; Harlen & Winter, 2004; Lynch, 2001), managerial and accountability purposes rely on marks and thus tests are the primary method of assessment (Delandshere, 2001; Shepard, 2000). These competing purposes are presented in the following sub-sections.

5.2.2.1. Preparation for the national examinations.

In Vietnam, in the final year in high school, Year 12 students are granted a diploma confirming that they have passed the General Education Diploma Examination (GEDE). The diploma is based solely on the marks that a student obtains in the GEDE. In addition to the GEDE, high school graduates have to pass a very competitive examination, the National University Entrance Examination (NUEE), in order to become a college or university student. The rates of students passing the GEDE and the NUEE influence schools' reputations. Therefore, preparing the students for these examinations was an essential part of classroom and school assessment in these schools.

High pass rates in the national examinations were important to the schools and students. According to Principal B, "[o]f course, high rates of students passing the GEDEs and the NUEEs create the reputation for the school." The students' passes in these examinations were considered to be the evidence as well as the target of learning. Principal B stated, "Eventually, students learn to take the national examinations, and they have to pass when they take the examinations. If you do not pass, skills or whatever you say become meaningless." These statements indicated that these examinations were high-stakes to the students.

The principals regarded the 45-minute tests and end-of-term tests as opportunities to prepare the students for these national examinations. According to Principal B, for end-of-term tests, "[t]he same test is given to all the students in the same grade at the same moment as a national examination so that the students can familiarise themselves with and prepare for the GEDE or the NUEE." In this way, the end-of-term tests followed the procedures used in the GEDE and the NUEE. In addition, through these school tests, the students were familiarised with the format of the national examinations. Principal A stated, "Through the tests, the students rehearse . . . . these are the rehearsals for the GEDE and the NUEE, to familiarise themselves with the format. That is why

in this school, English multiple-choice tests, for example, we follow the MOET's format." These excerpts revealed that preparing the students for these national examinations was a key purpose of assessment in the schools.

Because the principals considered the tests to be opportunities for the students to prepare for the national examinations, they expected the teachers to teach the students test-taking skills. Test-taking skills, also called test-taking strategies (Biçak, 2013) or test wiseness (Scruggs, White, & Bennion, 1986), refer to "the ability to use the format and characteristics of a test" as well as "the use of deduction and item cues to answer questions" in order to obtain high marks (Manly, Jacobs, Touradji, Small, & Stern, 2002, p. 342). Principal A expected his teachers to teach the students how to use such strategies. He said, "Show the students how to respond to test items. Respond quickly, even use tricks. In responding to test items, there are tricks, especially in multiple-choice items. Exclude this and this, for example. Do not think too much."

5.2.2.2. Professional purposes.

Some researchers use professional purposes to refer to the use of assessment by teachers and students to find out what students know and can do in order to modify and improve teaching and learning (Buhagiar, 2007; Curren, 1995; Gipps, 1994). The principals in this present study perceived that assessment informed the teachers about the students' learning and their own teaching so that they could be modified and improved. The teachers also thought that assessment informed students about their own learning so that they could improve their learning and achievement.

5.2.2.2.1. Informing the teachers for teaching modification and improvement.

The principals thought assessment helped the teachers determine their students' knowledge. Principal B pointed to this purpose.

> . . . before teaching a class at the beginning of a level, a new class, a new chapter, or a new lesson, the teachers must have the methods of assessment to get to know their students' level of knowledge and in what content areas they still have knowledge gaps so that they can support and provide knowledge prior to teaching, or during their teaching.

Principal B believed that this information helped the teachers form judgements about the students' knowledge in order to plan their teaching and facilitate the students' learning during instruction.

In addition, the principals thought that assessment informed the teachers about the students' learning. For example, Principal B believed that when embedded in instruction, assessment helped the teachers "see whether the students have understood, whether they have paid attention to their lesson." In the following criticism of the teachers' assessment of listening and speaking skills in

English instruction, Principal A suggested that regular assessment was necessary for the teachers to learn about their students' competence.

> . . . if you ask my teachers, "How are your students' listening and speaking?" they will intuitively say, "Weak", for example. They do not know about their students' competence in listening and speaking in detail because they do not often assess their students.

Principal A believed that frequent assessment of the students' listening and speaking skills provided the teachers with evidence about the students' competence in these skills and therefore reduced the possibility that the teachers would rely on their intuition rather than evidence when judging their students' competence.

The principals also believed that assessment informed the teachers about the effectiveness of their own teaching. "The teachers administer assessment to get feedback about their own teaching" (PA). Principal A considered assessment to be "a channel for the teachers to receive information about their own teaching and transferring knowledge." Similarly, Principal B believed that assessment could be embedded in instruction so that the teacher could evaluate his or her own teaching through the students' responses. He stated, ". . . continuous assessment is applied to small units of knowledge, when the teacher is teaching a unit of knowledge or presenting some content, in order to get responses from the students so that s/he can evaluate his/her own teaching . . . ." Both principals believed that through assessment the teachers evaluated the effectiveness of their own teaching through the students' feedback and responses to their teaching.

The principals believed that assessment helped the teachers modify and improve their teaching. Principal A perceived that assessment assisted in obtaining "information that helps the teachers modify their teaching." Principal B was more explicit about what the teachers relied on in order to improve their teaching. He said, "The teachers, through test results, high marks, low marks, improve their teaching career, their teaching methods." Principal B believed that the students' marks informed the teachers about the students' learning and encouraged them to improve teaching. The teachers' decisions about the modification of their teaching which relied on the students' marks may not be good decisions. Previous research has shown that in contexts where tests are high-stakes, teachers prepare their students for tests and raise marks but students' learning is not necessarily improved (Griffin, 2009). In order to know about students' learning, teachers' regular and systematic observations of students' learning and their dialogues and interactions with students are more important than marks (J. G. Brooks & M. G. Brooks, 1999).

5.2.2.2.2. Informing the students for learning modification and improvement.

The principals believed that assessment informed the students about their learning. Principal A thought that through assessment, the students "see how much knowledge they have learned from

their teachers, what percentage. . . . They check how much they have learned." Principal A's statement suggested that he assumed that assessment was used to measure the students' learning outcomes in order to inform the students about their own learning. Similarly, Principal B believed that the students relied on marks to know about their learning. He stated, "If students get 9s, they know that they satisfy the requirements in the unit of knowledge. If, for example, they get 4s, they know that they do not meet the requirement." The principals' views that marks informed the students about their learning suggested that they regarded assessment as a technology which was intended to measure the students' knowledge. When assessment relies on tests, which are used to measure students' learning outcomes for the purposes of placement, selection, or certification, it is considered to be a technology (Delandshere, 2001; Gipps, 1994). In this case, such qualities as standardisation, codification, efficiency, and reliability are valued. Therefore, assessment techniques and procedures are considered to be more important than how assessment is used to support and improve learning and teaching (Delandshere, 2001; Gipps, 1994). The principals saw tests as representative samples of knowledge which could be tested and then used to make generalisations about the students' learning outcomes. That is, they considered marks to be accurate measures of the students' learning outcomes and to have absolute meaning. However, in practice, in these schools marks were not accurate measures of the students' learning outcomes (to be discussed later in this chapter).

In addition to informing the students about their learning, the principals believed that assessment helped the students modify their learning. Principal A believed that "[f]or the students, assessment provides information that helps them modify their learning." Principal B was more specific about how test results or marks informed the students about their learning He believed their attitudes and behaviours also were changed. He said, "Test results modify the students' attitudes and behaviours in learning in order to meet the requirements and not to play . . . ." Principal B believed that marks changed the students' attitudes, behaviours, and even their commitment to learning.

The principals thought that through assessment, the students were provided with feedback that they could use to improve their learning. For instance, Principal B stated, ". . . the teachers help their students understand what the numbers mean to them, how they need to act, and how they must try hard to obtain marks." Principal B's statement suggested that he thought that the teachers' feedback could help the students improve their learning.

This sub-section indicates that the principals believed that assessment could help the teachers and students be informed about their teaching and learning so that they could be modified

and improved. However, the principals held a strong belief that marks informed both the teachers and students, who they believed relied on marks to modify and improve teaching and learning.

5.2.2.3. Managerial purposes.

The principals pointed to three purposes that were related to management: assessment for quality control, assessment for the school administrators' interventions, and assessment for reporting. Assessment was perceived as a means of quality control in the school. For instance, Principal B said, "When hearing this word (i.e., assessment), I think about a quality control system." When assessment is considered to be part of quality control, it is used to measure the final product (i.e., learning outcomes) for managerial and accountability purposes rather than to support and improve the process of learning (Middlehurst, 1997; Morley, 2003).

The principals also thought that assessment helped the school administrators make decisions about investment in the school facilities, professional development for the teachers, and collaboration with the students and parents in order to improve the students' learning.

Principal B stated, "The school administrators used test results to find ways to invest in facilities, provide professional development programmes for the teachers, educate the students in behaviours, and collaborate with parents to help the students learn well." Principal B's statement suggested that the school administrators relied on test results, or marks, in order to make decisions. Marks were also used as the basis for the school administrators to intervene in the schools' assessment practices. Principal A stated, ". . . through analysing marks, the school administrators take action towards the teachers and assessment practices."

The principals also perceived that assessment was administered to obtain marks for reporting to parents. According to Principal A, "[a]ssessment helps parents know about their children's competence." Similarly, Principal B stated, "Tests are administered to obtain marks to report to parents in order to inform them about their children's learning outcomes." The principals' statements suggested that marks were considered to be accurate measures of the students' learning outcomes, and they were used to inform parents about their children's learning.

This analysis indicates that the principals believed that assessment, specifically marks, were used as a means of quality control. In addition, they thought that the school administrators used marks as the basis for investment in facilities, for providing professional development to their teachers, for adjusting the students' attitudes and behaviours, and for collaborating with parents in order to improve the students' learning.

These managerial purposes may subordinate the professional purposes of assessment in these schools because when marks were considered to be measures of students' learning outcomes

and decisions rested on them, tests would focus on aspects of performance that could be marked as either correct or incorrect (Harlen, 2007). In other words, aspects of students' performance were narrowed. In addition, when schools and teachers were judged according to their students' marks, they would prepare their students for tests rather than using assessment as a means of improving learning (G. T. L. Brown & Harris, 2009; Harlen, 2007). The principals' accounts also suggested assessment had accountability purposes. These purposes are elaborated in the next sub-section.

5.2.2.4. Accountability purposes.

The principals identified two purposes related to accountability. They included controlling the students' learning and making them accountable and controlling the teachers' teaching and also making them accountable.

In these schools, the students' marks were used as a means of controlling the students' learning and making them accountable for their learning. Principal A believed that "[t]he students learn only if they are assessed. They only learn what is assessed; no assessment, they do not learn." Principal B was more explicit about what made the students learn. He said, "If there were not marks, the students would not try to learn. Therefore, there must be tests and marks." The principals' statements indicated that tests and test marks were believed to be extrinsic motivators for the students' learning. However, there is evidence that marks can motivate more competent students to learn (Stobart, 2008), but they have potentially negative influences on student learning in that they reduce students' motivation for learning, cause anxiety, and lead to narrow and surface learning (Harlen, 2007).

The principals perceived that marks were used to encourage the students to make an effort to learn. Principal A thought that tests should be challenging for the students because high marks made them complacent, and complacency prevented them from making an effort to learn. He stated, "For a test, it can be easy for the students to get 6, 7, or 8 marks, but it should be difficult to get 9 or 10 marks. If we give easy tests and most of the students get 9 or 10, then this does not encourage them to learn more."

In addition, the principals' statements suggested that the students' marks were used to control the teachers and make them accountable for their teaching. Specifically, the marks obtained by the students taught by different teachers were compared in order for the school administrators to make judgements about the individual teachers' teaching.

> I have directed that analyses of statistics be done carefully. At the end of a semester or a school year, there are statistics for each Section. I ask the Sections to compare (the results of the semester) with the results of the previous semester. Each Section has to show me how many classes Mr. A teaches and what the pass rates of his students are in comparison with

the Section's average pass rate. *If his pass rates are too high or too low compared with the Section's average, questions have to be raised.* (PA, emphasis added)

By comparing the marks obtained by the students taught by different teachers, Principal A controlled the teachers' teaching and made them accountable for their work. This practice could be considered problematic because it did not consider the nature of classes that might affect students' test results. The principals' judgements about the individual teachers' teaching based on the students' marks may push the teachers to manipulate the students' marks. This could be the reason for the teachers' attempt at constructing the 15-minute tests according to their students' competence (to be discussed later in this chapter).

In short, the principals considered assessment, especially marks, to serve accountability purposes. Specifically, marks were used to control learning and teaching and make the students and teachers accountable for their work.

The analysis has shown that these principals recognised multiple purposes of assessment. The tests, especially the 45-minute and end-of-term tests, were used as opportunities to prepare the students for the national examinations. Marks were thought of as accurate measures of the students' learning outcomes and used for professional purposes (i.e., informing and improving teaching and learning), for managerial purposes (i.e., quality control, making decisions about investment in facilities, providing the teachers with professional development, adjusting the students attitudes and behaviours, collaborating with parents to improve the students' learning). Marks were also used for accountability purposes (i.e., controlling learning and teaching and making the students and the teachers accountable for their work). These purposes competed with each other. When assessment is administered to obtain marks for managerial and accountability purposes, the accuracy of marks as measures of students' learning outcomes is prioritised because marks are high-stakes to both students and teachers (Harlen, 2007). In addition, when assessment is administered to obtain marks for managerial and accountability purposes, tests, especially multiple-choice tests, are the primary assessment method (Delandshere, 2001). However, when assessment is administered for learning improvement, students take an active role in the assessment process (Black & Wiliam, 1998a; Estaji, 2011). Furthermore, they are assessed mainly through alternative assessments which are embedded in instruction (Black & Wiliam, 1998a; Harlen & Winter, 2004). Not only the product of learning (i.e., learning outcomes) but its processes are of interest to the teachers because students improve learning through using teachers' and peers' descriptive feedback (Black & Wiliam, 1998a; Stobart, 2008).

This study has shown that the managerial and accountability purposes may subordinate the professional purposes in the context of these. For example, although the principals perceived that

assessment had multiple purposes, the managerial and accountability purposes were prioritised because in these schools, "[a]t present our assessment means marks" (PA) and "[t]he final information that exists in educational assessment is the measures, marks" (PB). The principals' concerns about marks were manifested in their concerns about technical and ethical issues associated with assessment in these schools. These concerns are elaborated in the next section.

5.3. The Principals' Concerns about Technical and Ethical Issues

The principals' accounts in the interviews showed that they were concerned about the technical issues of assessment (i.e., tools and procedures), the ethical issues in test construction, and objectivity and fairness in assessment administration. These concerns are discussed in the following sub-sections.

5.3.1. Concerns about technical issues.

The technical aspects of assessment include standardisation, validity, and reliability (Gipps, 1994). When assessment is assumed to be a technology that is used to measure learning outcomes, emphases are put on methods and procedures of assessment rather than on knowing about individual performance and learning (Delandshere, 2001). The principals assumed tests to be a technology which could be used to accurately measure learning outcomes. As a consequence of this assumption, they were concerned about technical issues of assessment such as the coverage of knowledge in the tests, the accuracy of the students' marks (i.e., marks were considered to be accurate measures of learning outcomes), standards of knowledge and skills required by the MOET, and procedures for achieving accuracy.

5.3.1.1. Tests as a technology for measuring knowledge.

The principals' accounts suggested that they considered tests as a technology used to measure the students' learning outcomes. For example, Principal A equated the students' ability to do tests with an indication of their learning. He said, "When I have a test bank of, say, 30 tests, if my students can do all these 30 tests, it is good. It means they have learned well." This statement suggested that Principal A assumed that knowledge could be assessed and judged by using tests alone. Indeed, the MOET regulated that the EFL teachers in high schools across Vietnam administer two oral tests, two 15-minute tests, two 45-minute tests, and one end-of-term test in each semester, and that the judgements about students' learning outcomes are made in the form of marks (MOET, 2009e, 2011f).

The principals' perceptions of the coverage of knowledge in the tests further indicated that they assumed that tests could be used to measure the product of learning (i.e., students' knowledge).

Assuming that knowledge could be organised into distinct parts and measured, Principal B believed that periodic tests should have broader coverage of knowledge than that of continuous tests. He said, "As for periodic tests, the coverage of knowledge has to be broader because the tests are administered at the end of a unit, a chapter, or a semester in order to summarise a period of learning." In a similar vein, Principal A thought that the more knowledge was included in the tests, the more accurate the judgements about the students' learning. He stated,

> Tests, especially in English, have to cover the curriculum, not just some areas. Tests that cover only some areas of the curriculum do not help make accurate judgements about the students' knowledge; they may have good luck. Tests have to cover all the curriculum to make accurate judgements about the students' learning.

The principals considered assessment to be a technology that could offer accurate measures of the students' learning outcomes. For instance, Principal B expected that the marks that the students obtained from the tests were accurate measures of their learning outcomes. He thought that "the goal of assessment is to get accurate information." Similarly, Principal A said, "Marks must be the accurate measures of the students' learning outcomes."

This sub-section indicates that the principals considered tests to be a technology that could be used to measure the students' knowledge accurately. In addition, their statements suggested that they regarded tests as the primary tools used to measure the students' knowledge, most often at the end of a learning unit. The principals' perceptions of assessment as a technology suggested that they were concerned about the product of learning and the quantitative information about learning instead of the processes of learning and the qualitative and holistic information about the students' learning. Considering tests as a technology used to obtain marks, the principals were concerned about tests and the procedures for test administration. These concerns are presented in the following section.

5.3.1.2. Concerns about the tests and the procedures for test administration.

Because the principals assumed that tests could be used to measure the students' knowledge, they were concerned about the tests themselves and the procedures for test administration. Specifically, they were concerned about the standards (i.e., the MOET's requirements about the minimum competence in the knowledge and skills that the students had to achieve at the conclusion of a learning unit or a grade) and the procedures for test administration.

Being concerned about the standards of knowledge and skills as stated by the MOET, Principal B thought that "test questions and levels of knowledge have to follow the standards." Principal B's concern may be regarded as a concern about the technical issues of assessment.

These principals were also concerned about the procedures for test administration to ensure that the students' marks reflected their learning outcomes. Principal B stated,

> Only when tests are well administered do the school administrators and the teachers have the condition, the reliable information, to reform teaching and learning methods and continuously improving the quality of teaching and learning as expected by the MOET, by society, parents, and students.

In this statement, Principal B indicated that the careful application of procedures for test administration would ensure the quality of the information about the students' learning outcomes obtained. However, no matter how the tests in these schools were administered, the information about the students' learning outcomes obtained from these tests was not comprehensive because these tests represented only snapshots of the knowledge that the students had been taught (Harlen, 2007). Indeed, other researchers have argued that factors such as the reliability of the tests, test anxiety, and cheating also impact on the accuracy of the information gathered from tests (Estaji, 2011; Must & Must, 2013).

The analysis revealed that these principals focused on technical aspects of assessment. They were concerned about the coverage of knowledge in the tests, the standards of knowledge against which the tests measured the students' knowledge, the accuracy of the students' marks as measures of their learning outcomes, and the procedures that could ensure that marks were accurate measures of the students' learning outcomes. In addition to these concerns, the principals were concerned about the ethical issues associated with the teachers' test construction, objectivity, and fairness in assessment. The next sub-section discusses these ethical concerns.

5.3.2. Concerns about ethical issues.

The words "moral" and "ethical" can be used interchangeably (Davies, 1997). In assessment, ethical issues are related to test developers' competence; the examinees' right to privacy and confidentiality associated with reporting test scores and students' rankings; support or violation of democratic values such as equality, liberty, and justice (Delandshere, 2001); and bias and fairness (Messick, 1996). In addition, students' cheating and teachers' deliberate preparation of their students to respond to test items selected for important tests and examinations as well as teachers' abuse of the power associated with assessment for reward and punishment are related to ethical issues in assessment (Bandaranayake, 2011; Estaji, 2011).

The principals' accounts indicated that they were concerned about some ethical issues in assessment. Specifically, they were concerned about the teachers' test construction and their objectivity and fairness in test administration. The following sub-sections elaborate these concerns.

5.3.2.1. Ethical issues associated with the teachers' test construction.

In Vietnam students are ranked according to the MOET's criteria; they are not ranked against each other. Specifically, students are ranked as very poor, poor, average, good, or outstanding students according to the average mark they obtained in a semester and/or a school year (MOET, 2011f). Therefore, to be fair to all the students, the teachers should construct the tests using the same standards (i.e., the MOET's standards of knowledge and skills) because judgements about students' learning outcomes are morally legitimate when they are based on acceptable standards (Curren, 1995). However, the principals believed that the teachers in these schools constructed tests according to their own expectations of the students, the students' competence, and the marks of students from previous years rather than according to the MOET's standards of knowledge and skills.

According to Principal A, "the teachers have different expectations when designing the 15-minute tests." In addition, ". . . the expectations in assessment in this school are quite high. . . . the teachers in this school require more of their students" (PA). These practices were not ethically legitimate because 15-minute tests in School A had different levels of difficulty and these tests were more difficult than those in other schools. As a consequence of this, the students in different classes may have obtained different marks not because of their competences but because the tests they took had different levels of difficulty. This practice violated the "equal treatment" principle (Elwood, 2013, p. 207), which suggests that students are provided with equitable conditions when taking a test. Consequently, different marks meant they were assigned different rankings. This was an unfair practice.

The principals believed that their teachers also constructed tests according to their students' competence. In School A, where the "more able" 9th Graders in the city enrolled, the teachers constructed tests that the principal perceived to be "difficult" tests.

> Depending on the students' competence, the teachers will, in their tests, use high-level application questions. For example, other schools do not give tests with high-level application questions; only knowledge, comprehension, and low-level application questions are included. In this school, the teachers give tests with high-level application questions. . . . We go beyond the MOET's standards. We use high-level application questions and require a little more of our students. (PA)

In describing the teachers who created tests with high-level application questions, Principal A referred to his teachers of all subjects, including the EFL teachers.

In School B, which enrolled the 9th Graders who did not qualify to enroll in the public schools in the city, the teachers gave their students tests of "moderate" difficulty. "Because our students are those who could not be admitted to the public schools in the city, the teachers give them moderate tests" (PB). The principals also believed their teachers constructed both continuous

and periodic tests according to their own students' competence. They believed that these teachers did not assess their students according to the MOET's standards. Therefore, they were concerned about their teachers' test construction. The principals' concern was understandable. Tests that were constructed according to students' competence might be good for teachers to find out their students' knowledge gaps and provide feedback for the students to improve their learning. However, when these tests were administered to generate marks for ranking students, more able students were disadvantaged because they were given more difficult tests.

The principals also thought that their teachers constructed continuous and periodic tests according to the marks of the previous year's students.

> The Sections have to calculate statistics for their own Section. The Sections use the statistics to consider how last year's 10th and 11th Graders were. Higher? Lower? They compare with last year's tests. For example, last year's tests were like this, and this year's students are weaker or better, then they lower or raise the level of difficulty a little bit. That is, they use the statistics as the basis for adjusting the content and amount of knowledge in the tests. (PA)

The teachers' use of marks as a basis for constructing tests indicated that they did not construct the tests according to the MOET's standards. This was not a morally legitimate practice.

This sub-section indicates that the principals believed that the teachers in these schools did not apply the MOET's standards of knowledge and skills in their tests. They thought that the teachers constructed tests according to their own expectations of the students, according to the students' competence, and according to the marks obtained by the students in the school in the previous year. The marks obtained from these tests, therefore, were not accurate measures of the students' learning outcomes. The teachers' test construction challenged the principals' perceptions that tests could be used to accurately measure the students' learning outcomes. It also challenged the principals' perceptions of the purposes of assessment. When the marks were not accurate measures of the students' learning outcomes based on the MOET's standards, the judgements that the principals made about teaching and learning in their schools were not accurate. These judgements were unfair to the students and teachers. In addition, the school administrators' decisions about investment in facilities and collaboration with the teachers, parents, and students in order to improve the students' learning were based on the students' marks, while the marks did not reflect the students' learning outcomes. The ethical issues related to the teachers' test construction were the cause of the principals' concerns about objectivity and fairness in assessment in their schools. The next sub-section elaborates these concerns.

5.3.2.2. Concerns about objectivity and fairness in assessment.

Objectivity and fairness in assessment have attracted the attention of researchers and many writers in assessment. A common view of objectivity "stresses the importance of eliminating prior assumptions, past experiences, and personal convictions" (Bulterman-Bos, Terwel, Verloop, & Wardekker, 2002, p. 1087). According to this traditional view, teachers' personal interpretations, opinions, experiences, and judgements should be minimised when they administer assessment (Bulterman-Bos et al., 2002).

There have been many definitions of fairness (see Camilli, 2013; Gipps & Stobart, 2009; Kane, 2010; Tim McNamara & Ryan, 2011; Sireci & Rios, 2013; Xi, 2010). However, researchers agree that fairness includes the absence of bias (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Kunnan, 2004, 2010). Bias "refers to construct-irrelevant components that result in systematically lower or higher scores for identifiable groups of examinees" (American Educational Research Association et al., 1999, p. 76). In addition, fairness is achieved when students have equity in their opportunity to learn the content covered in the test (American Educational Research Association et al., 1999; Kunnan, 2004). Fairness is also achieved when all students receive equitable treatment in the testing process or they take a test in the same conditions (American Educational Research Association et al., 1999; Baharloo, 2013).

Some researchers argue that fairness needs to be considered from two different perspectives, depending on the primary purpose of assessment (i.e., measuring students' knowledge and skills or supporting and improving learning) (Baharloo, 2013; Lantolf & Poehner, 2013). From the traditional view, "fairness requires equitable treatment of all test takers in the testing process" (Baharloo, 2013, p. 1931). This view is applied when assessment is administered to measure students' learning outcomes for admission, accountability, selection, and promotion (Baharloo, 2013). When assessment is conducted mainly to support and promote learning, fairness has a different meaning. In this case, "fairness requires providing each individual with what they deserve regarding their needs" (Baharloo, 2013, p. 1935). Specifically, students receive support from peers, teachers, and materials, and "each individual receives as much assistance as he or she needs" (Baharloo, 2013, p. 1935). Standardisation is prioritised in traditional assessment because equal treatment is required to ensure fairness (Lantolf & Poehner, 2013). When assessment is for learning, "[f]airness requires doing everything possible to maximally support individual learner development, with the understanding that some individuals will need more time and resources than others" (Poehner, 2011, p. 103).

The principals' accounts suggested that they valued objectivity and fairness in assessment. For example, Principal B said, "I develop in the teachers a sense of objectivity and fairness." He required his teachers to "be objective and fair to all the students" in assessment.

The principals reported a variety of reasons for their concerns about objectivity and fairness. First, there was a concern about the inconsistency among the teachers in marking the students' test papers. Principal A said, "For the same error, some teachers give one mark, some give half a mark, and some do not give any mark. This creates big differences." Second, there was a concern about unfairness created by some teachers offering extra lessons in their homes. Principal A stated, "The teachers teach extra lessons and they may use their tests to teach their students in the extra lessons. Those students who attend these teachers' extra lessons are more advantaged." In Vietnam, many school teachers teach extra lessons in their homes for additional income, and some of them are said to construct tests that favour those students who take their extra lessons. This practice made Principal A feel concerned about the teachers' objectivity and fairness in assessment in his school.

Because the principals valued objectivity and fairness, they proposed a number of measures that they thought could help achieve these qualities. These measures were intended to provide all the students with the same conditions in the testing process. For example, the teachers were not allowed to construct periodic tests for their own students. School A "regulates that 45-minute and end-of-term tests are drawn from the school's test bank." Similarly, in School B, "the school administrators do not allow the teachers to assess their own classes" in periodic assessment (PB).

Additionally, the principals believed that equitable testing conditions were necessary for objectivity and fairness. Specifically, they required the teachers to follow mandatory procedures for test administration and marking. In both schools, all the students in the same grade were given the same end-of-term test, and "[t]he purpose of using the same tests for all the students in a block is to ensure objectivity, and to avoid tests with different levels of difficulty" (PA). In order to ensure that all students were treated equally during testing, test administration was standardised. For instance, for end-of-term tests, all the students in the same grade in School A took the same test "at the same hour, on the same day" (PA). In addition, the teachers had to mark the students' test papers without knowing the names of the students. For example, in School B, "[i]n order to ensure objectivity, test papers are marked anonymously" (PB). Anonymous marking was intended to minimise the potential influence of the teachers' prior knowledge about their students on the marking. Also, the marks obtained by the students in different classes were compared in order to make sure that the students taught by different teachers obtained similar marks. For instance, in School A,

> . . . the school administrators check the score books to see if there is anything unusual. For example, why are these classes different? Or among different teachers, we make sure that the marks given by different teachers are not too different. . . . Or for the same teacher, are

the test results of this class and those of that class too different? The purpose is to make sure that assessment has to be fair and objective. (PA)

The above statement suggested that Principal A assumed that objectivity and fairness were achieved when the students in the same class and across classes obtained comparable marks. This interpretation of fairness is "popular" but is incorrect because "[s]core differences between demographic groups do not necessarily mean that a test is unfair" (Zieky, 2013, p. 294).

The principals were more concerned about objectivity and fairness than about the professional purposes of assessment. For instance, Principal A was well aware that using the tests from the school's test bank deprived the teachers of opportunities to administer assessment for learning, and he considered this practice "a contextual solution."

> Indeed, using the tests from the school's test bank is not good practice. It is not good for several reasons. For example, I teach this class and I design a test to assess some content according to my intention. Or when assessment is carried out for the purpose of practicing, I will assess the content which my students are weak on. Right? But when I draw a test from the school's test bank, this test may assess a different content area.
> However, this is our choice, frankly speaking, because of the teachers' extra lessons in their homes for extra incomes. It is not good. . . . the teachers are deprived of their autonomy. But it is a contextual solution. (PA)

Principal A recognised the trade-off or compromise when the teachers had to use the school's 45-minute and end-of-term tests instead of constructing their own tests. However, because some of the teachers provided extra lessons in their homes, the principals thought that they may not have treated all of the students fairly. Therefore, the purpose of supporting students' learning was subordinated to the purpose of achieving objectivity and fairness in assessment.

The discussion in this section indicates that the principals emphasised objectivity and fairness in assessment in their schools. They believed that the teachers' inconsistency in marking and their extra lessons may have influenced their objectivity and the desire for fairness in assessment. Consequently, they proposed procedures that the teachers had to follow to ensure objectivity and fairness. The principals thought that achieving objectivity and fairness in assessment was more important than using assessment for the purpose of improving learning. The principals' emphasis on objectivity and fairness was also related to their expectations of the teachers' compliance with the regulations and procedures for assessment and their distrust of the teachers. The next section focuses on these issues.

5.4. The Principals' Expectations of the Teachers' Compliance and Their Distrust of the Teachers

The principals expected their teachers to comply with the regulations and procedures for test administration in the schools. The principals' expectations of the teachers' compliance were

associated with their concerns about the technical and ethical issues associated with assessment (discussed in the previous section) and their lack of trust in the teachers' professional competence, commitment, and fairness. The following sub-sections elaborate the principals' expectations of the teachers' compliance and their distrust of the teachers.

5.4.1. The principals' expectations of the teachers' compliance.

According to the principals' accounts, assessment in these schools was regulated by the MOET and the DOET, so the school staff had to comply with these regulations. In addition, the school administrators had their own regulations and procedures surrounding assessment related to the particular schools.

5.4.1.1. Expecting the teachers to comply with the MOET's and the DOET's regulations.

The principals' reports indicated that assessment in these schools was controlled by the MOET's and the DOET's regulations. Principal B said that assessment in his school was "administered according to the procedures and regulations of the MOET" and "[t]he school staff cannot administer assessment in other ways; the school staff can only administer tests with small variations." These statements indicated that staff had very limited autonomy in administering assessment in their schools.

There were regulations related to the administration and timing of tests. Principal A stated,

> In this school, assessment is administered according to the MOET's regulations, with two forms of tests, continuous tests and periodic tests. Specifically, there are oral tests, 15-minute tests, 45-minute tests and end-of-term tests. Generally speaking, the timing and the forms of tests follow the MOET's regulations.

The principals reported that the school staff also had to comply with the MOET's and the DOET's regulations related to the type of test items. Principal B said, "We have to comply with the MOET's and the DOET's guidelines: periodic tests consist of 100% multiple-choice questions." Also, the schools had to administer the number of tests regulated by the MOET. Principal A said, "[A]ccording to the MOET's regulation, each subject has a specific number of tests."

Complying with the MOET's and the DOET's regulations about assessment was very important to the principals. Principal A said, "I always direct that the MOET's and the DOET's regulations be strictly implemented. . . ." This statement suggested that implementing assessment as regulated by the MOET and the DOET was prioritised and the teachers' autonomy in assessment was second to their compliance with the regulations.

5.4.1.2. Expecting the teachers to comply with the schools' regulations and procedures.

Apart from requiring the teachers to comply with the MOET's and the DOET's regulations, the principals and school administrators required them to follow the schools' regulations and procedures for assessment. Through these, the principals controlled the teachers in test construction, test structure, and test administration. In test construction, "[t]he school regulates that the teachers prepare a test matrix following the requirements of the lesson, the subject, and the form of assessment, and design a test using the test matrix" (PB). When constructing the tests, the teachers had to follow the principals' decision about the proportion of marks allotted to the different items in a test. "I direct that the part for application questions takes up 1 to 2 marks. . ." (PA). In test administration, "[t]he teachers have to carry out tests at fixed times. After two weeks, the teachers have to hand in the statistics related to the marks" (PA). In general, "the school has procedures which have been disseminated, and the teachers must comply with the school's testing procedures" (PB). This compliance limited the teachers' autonomy to construct and administer assessments for the purposes of improving student learning.

Because the principals required their teachers to comply with the regulations and procedures for assessment, they equated the teachers' knowledge and skills in assessment with the teachers' fulfilment of these regulations and procedures. In other words, the principals regarded the teachers' administration of assessment as per the regulations as signs of knowledge of and skills in assessment. For instance, Principal B said, "Basically, the knowledge and skills in assessment of the teachers in this school meet the requirements and criteria that the MOET proposes . . . ." Similarly, Principal A said, "I think that the EFL teachers' skills in assessment are quite good. They administer assessment quite well according to the regulations." These remarks suggested that the teachers' compliance with the regulations and the mandated procedures for assessment were equated with their skills in assessment.

The principals believed that the teachers only did what they were required to do. In other words, "[t]hey only implement the regulations" (PA). Principal A believed that one reason, among others, why the teachers did not assess their students' listening skills was that there was no regulation related to assessing these skills. He believed that they did not assess listening because "there is no specific regulation about assessing listening. There is no regulation about when and how to assess listening" (PA). He also stated that the teachers did not correct the errors in the students' test papers, and he believed that "[i]f there is a regulation, the teachers will correct their students' errors" (PA). In a context where compliance was valued and required, avoiding the activities that were not mandated may have been a safe choice for the teachers. (The reasons for the teachers' compliance are discussed in the next chapter).

This sub-section indicates that the principals expected the teachers to comply with the MOET's, the DOET's, their own, and their school administrators' assessment regulations and procedures. These regulations and procedures influenced the teachers' autonomy in the construction and administration of tests. The principals' expectations of the teachers' compliance was linked to their lack of trust in the teachers. The next section discusses the principals' distrust of the teachers.

5.4.2. The principals' distrust of the teachers.

The principals' reports indicated that they distrusted the teachers' professional competence, commitment, and fairness. First, the principals thought that the teachers' professional competence was limited. Specifically, Principal A thought that his teachers could not distinguish between knowledge questions, comprehension questions, and application questions.

> At present, it is up to the test designer (to consider a test question to be a knowledge, comprehension, or application question). Frankly speaking, it is a reality. . . . For example, in designing the end-of-term tests, such a situation occurred. . . .
> In semester 1, a teacher assumed one question to be a high-level application question, but at a Section meeting, it was assumed to be a low-level application question. Or one question may be assumed to be a knowledge question, a comprehension question, or an application question by different teachers. . . . At present, what knowledge means, what comprehension means, and what application means are still ambiguous. (PA)

The above extract indicated that Principal A doubted his teachers' competence in designing test items with different cognitive levels as they were required to do when constructing tests according to the test matrix (i.e., a table which showed the parts of a test, the number of test items in each part, the types of the test items, and the proportion of marks for each test item).

In addition, Principal A thought that his teachers had inadequate skills for designing multiple-choice test items. He believed that his teachers were "still inexperienced in designing multiple-choice tests." He remarked, "The teachers are very weak in designing multiple-choice test items with quality distractors. I can say that the teachers in all subjects are weak in this task. I am sure."

The principals also doubted the teachers' competence in administering assessment. For example, Principal B thought that his teachers were not competent enough to use assessment information to identify the students' weaknesses. He said, "My teachers still have limitations in using tests to identify the students' weaknesses." Additionally, Principal A expressed his distrust of the teachers' skills in speaking English and listening to it. He stated, "I think, first, frankly speaking, the teachers do not assess speaking and listening because their own speaking and listening skills are not good. Their own speaking and listening skills are not very good."

Second, the principals expressed distrust of the teachers' professional commitment to their work. For example, Principal A said that his teachers "do not correct the students' errors in the test

papers carefully and responsibly" and "[t]hey even do not synthesise these errors." According to Principal A, among other reasons, "the main reason is that they are lazy."

Third, the principals also did not believe their teachers could be fair in assessment administration. This distrust manifested itself in test administration in both schools. In both schools, the 45-minute and end-of-term tests were kept under secure conditions by the principals, and in School A the tests to be used were selected by the vice-principal on the day of test administration. For example, in School A, "the 45-minute and end-of-term tests are not to be known by all the teachers in the Section because of secrecy" (PA). In School A, with respect to tests with higher weightings (i.e., periodic tests), the vice-principal "chooses a test from the school's test bank. He is the only person who knows which test is given to a class" (PA). The vice-principal also photocopied these tests and answer sheets, and he gave the test papers and answer sheets to the teachers just before the teachers administered them to their students. The procedures used to keep the periodic tests secret from the teachers indicated that the school administrators did not trust the teachers to be fair when administering the tests. Together with the principals' expectation that the teachers comply with the assessment regulations and procedures, the principals' distrust of the teachers' fairness limited the teachers' autonomy in assessment.

This section shows that the principals distrusted the teachers' competence, commitment, and their ability to be fair in assessment. The distrust manifested itself in the principals' beliefs about the teachers' incompetence in identifying the cognitive levels of test items, in designing quality multiple-choice tests items, and their command of English. The principals also distrusted the teachers' commitment to giving feedback to the students and in the teachers' fair treatment to all the students in assessment. The distrust was associated with the principals' expectations of the teachers' compliance with the regulations and procedures for assessment.

5.5. Discussion

The interviews with the principals were analysed to know about their perceptions of assessment in their contexts. There were five major findings. First, the principals perceived that assessment had multiple purposes. They valued assessment in teaching and learning and indicated four purposes of assessment. They reported that through assessment, they prepared their students for the national examinations. This finding was congruent with the EFL teachers' perception that students needed to be familiar with the knowledge and skills tested in the national examinations and the format of these examinations. In Vietnam, 12th Graders have to pass the GEDE in order to get a certificate indicating that they have finished their high school education. In order to study at a college or university, students who have passed the GEDE sit the NUEE, which is very competitive.

Therefore, these two national examinations were extremely important to high school students and their parents. Similar to the principals in the present study, the principals in a study in Hong Kong believed that preparing students for high-stakes tests was necessary (Ngan et al., 2010). The principals in the Hong Kong study thought that students needed to be familiar with the format of high-stakes tests. In another study in Western Canada, principals also thought that they needed to prepare their students for large-scale assessment (Newton et al., 2010). The findings in the present study and those in studies carried out in various contexts suggested that exam preparation is part of classroom assessment in contexts where testing is high-stakes.

The principals in the present study believed that assessment informed teachers and students and helped them modify and improve teaching and learning. These perceptions about the formative purposes of assessment echoed policy-makers' expectations of reform in assessment at the school level (see Chapter 4). They were also similar to the EFL teachers and the students' perceptions about the purposes of assessment (see Chapter 5 and 6). Principals in other contexts such as Western Canada (Newton et al., 2010) and Michigan, the USA (Shen et al., 2010) held similar perceptions. For example, the principals in Michigan believed that assessment informed students, teachers, and principals about student learning and helped teachers improve their teaching and student learning (Shen et al., 2010). The principals in Western Canada also believed that assessment data informed teachers about areas of the curriculum that they needed to improve (Newton et al., 2010). Similar to the principals in the study in Michigan (Shen et al., 2010), the principals in the present study believed that assessment informed parents about their children's learning.

The principals in the present study perceived that assessment data helped them change teaching and assessment practice for improvement. Principals in various contexts such as Florida, the USA (Jones & Egley, 2006), Michigan (Shen et al., 2010), and Western Canada (Newton et al., 2010) reported that they used assessment data for school-based planning and priority-setting in order to improve teaching and learning in their schools.

The principals in the present study thought that assessment was used to control teaching and learning and to make teachers and students accountable for their work. This finding was supported by one of the perceptions about the purposes of assessment held by the EFL teachers, who thought that assessment made themselves and their students accountable for their work (see Chapter 6). This perception was also held by principals in other contexts (Newton et al., 2010; Shen et al., 2010). The principals in these studies perceived accountability to be the most dominant purpose of assessment.

Second, the principals in the present study were concerned about technical issues in testing. They assumed that tests could accurately measure student learning outcomes. In addition, they

believed that tests needed to consist of a broad coverage of knowledge and to be designed according to the MOET's standards of knowledge and skills. However, the teachers in this study reported that they designed tests according to their expectation of their own students and according to the students' competence. In addition, the teachers reported that they modified the difficulty of their tests in order to raise their own students' marks. This suggested that the principals' expectation about the standardisation of tests in their schools were not fulfilled by the teachers, who perceived that students' marks had consequences for their students and for themselves. The principals were also concerned about the accuracy of marks as measures of learning outcomes. However, the EFL teachers' test construction and modification indicated that marks were not accurate measures of students' learning outcomes. In addition, the students perceived that marks did not reflect their learning outcomes because some students cheated to obtain high marks. Similar to these concerns of the principals in the present study, principals' concerns about technical issues in assessment were a finding in Newton, et al.'s (2010) study. The principals in Newton and colleagues' study were concerned about reliability and validity in assessment in their schools.

Third, the principals in the present study were concerned about their teachers' objectivity and the ability to be fair in assessment. They believed that the teachers constructed continuous tests according to their own expectations of their students and the students' competence rather than according to the MOET's standards. These beliefs were echoed by the EFL teachers' report that they designed 15-minute tests according to their expectation and the students' competence (see Chapter 6). The principals valued objectivity and fairness in assessment. They thought that there was a lack of fairness because some teachers were inconsistent in marking and some other teachers constructed continuous tests that advantaged those students who took private extra lessons in the teachers' homes. The students in the study also believed that some teachers designed tests that advantaged those students who took private lessons in the teachers' homes. The principals reported that they tried to ensure objectivity and fairness in testing by giving the same periodic test to all the students in the same grade and asking the teachers to mark these test papers anonymously. The principals thought that ensuring objectivity and fairness in testing was more important than giving the teachers autonomy to use tests for teaching and learning improvement. The principals' concern about objectivity and fairness reflected one of the expectations of policy-makers (i.e., reform in assessment helped ensure accuracy, objectivity, and fairness) (see Chapter 4). However, objectivity and fairness were not achieved in these schools because the EFL teachers reported that they used marks to engage their students in classroom activities (see Chapter 6) and they used marks as rewards and punishments (see Chapter 6 and 8).

Fourth, the principals expected their teachers to comply with the assessment regulations and procedures. The principals expected the teachers to comply with the MOET's, the DOET's, and the schools' regulations and procedures for assessment. This finding was congruent with the EFL teachers' report that the school administrators controlled their test format, the timing of test administration, test administration, and marking (see Chapter 6). Both the principals and the teachers perceived that the principals' expectation of the teachers' compliance and the teachers' compliance restricted the teachers' autonomy in assessment, especially assessment for formative purposes. This suggested that in these schools, assessment for summative purposes rather than assessment for formative purposes was prioritised.

Fifth, the principals distrusted their teachers in assessment. Specifically, they thought that their teachers lacked knowledge and skills in assessment. This finding was similar to a finding from the analysis of the policy and professional development documents (Chapter 4). The principals thought that their teachers had difficulty distinguishing knowledge questions, comprehension questions, and application questions. They also perceived that their teachers were inexperienced in designing quality multiple-choice test items. The principals believed that their teachers had difficulty in using assessment to identify the students' difficulties in learning. Similar to the principals in the present study, the principals in Newton, at al's (2010) study thought that their teachers were not well prepared for interpreting and using assessment data for teaching and learning improvement. The principals in the present study also believed that their teachers did not give quality feedback to help the students improve their learning. These beliefs were based on a view that the teachers were not committed to their work. This belief was supported by the feedback, mostly direct corrective feedback and praises about the students as persons, that the teachers gave their students in the observed classrooms. The principals' distrust of the teachers was also manifest in their thought that their teachers were not fair when they constructed tests and marked the students' test papers. This distrust coincided with the students' perception that their teachers were not fair when constructing and marking tests (see Chapter 7). It appeared that the high stakes or consequences attached to students' marks did more harm than good. They caused distrust among teachers (e.g., some of the teachers in this study thought that other teachers gave their students easier tests because they wanted their students to obtain good marks) (see Chapter 6). They also caused the principals' and the students' distrust of the teachers.

This chapter reflects a general picture of the contexts of assessment in the schools from the principals' perspectives. The next chapter will bring about more insights into the contexts of assessment in these two schools from the teachers' perspectives.

.

Chapter 6. The Context of Assessment as Perceived by the EFL Teachers

6.1. Introduction

This chapter discusses the results of the analysis of the interviews with six EFL teachers. The chapter is organised according to the themes and theme clusters in Figure 3. The first cluster is related to the teachers' perceptions of the purposes of assessment. The teachers perceived that assessment had five purposes. First, they thought that assessment helped prepare their students for the national examinations. Second, they perceived that assessment helped them modify their teaching and help students improve their learning. Third, the teachers thought that assessment made teachers and students accountable for their work. Fourth, they thought that assessment was implemented to summarise students' learning outcomes for ranking students and reporting to parents. Fifth, the teachers reported using assessment to control their students' learning.

The second cluster is about the teachers' perceptions of marks. The first theme indicates that the teachers assumed assessment as a technology for measuring students' knowledge and generating marks. The second theme is related to the teachers' perceptions of the MOET's, their school administrators', students', parents', and their own expectations and concerns about marks.

The third cluster comprises the teachers' perceptions of the school administrators' control over their assessment and the reasons for their compliance. The first theme reveals what the school administrators control, and the second theme is about the reasons for the teachers' compliance.

The fourth cluster is related to the teachers' perceptions of accuracy, objectivity, and fairness in assessment. The first theme indicates that the teachers valued these qualities. The second theme shows ethical issues related to the teachers' test construction.

Apart from the four sections that discuss the four theme clusters, the last section discusses the findings of the chapter.

Figure 3. Results of the Analysis of the Interviews with the EFL Teachers

The Context of Assessment as Perceived by the EFL Teachers

Perceived Purposes of Assessment	Marks as the Main Concerns	Control and Compliance	Expectations of Accuracy, Objectivity, and Fairness	Clusters

Preparing the students for the national examinations
- An important part of the schools' activities
- Familiarising the students with knowledge and skills tested in the national examinations
- Familiarising the students with the format of the national examinations

Modifying and improving teaching and learning
- Informing the teachers about their own teaching
- Helping the teachers modify teaching
- Informing the students about their learning
- Helping the students improve learning

Making the teachers and students accountable for their work

Summarising learning outcomes for ranking and reporting

Controlling the students' learning
- Using marks to engage the students in classroom activities
- Using marks as punishment

Assessment as a technology for measuring knowledge and obtaining marks
- Equating assessment with measuring knowledge
- Equating assessment with assigning marks

The teachers' perceptions of the stakeholders' expectations and concerns.
- The teachers' perceptions of the MOET's expectations
- The teachers' perceptions of the principals' expectations
- The teachers' perceptions of parents' concerns about marks
- The teachers' perceptions of the students' concerns about marks
- The teachers' expectations of the students' marks

The school administrators' control of the teachers' administration of assessment
- Controlling test format
- Controlling the timing of test administration
- Controlling test construction
- Controlling test administration
- Controlling marking

The teachers' acceptance of compliance and their justifications
- Compliance as a social norm
- Compliance for job security

Valuing accuracy, objectivity, and fairness
- Valuing these qualities
- Supporting measures intended to ensure these qualities

Ethical issues related to the teachers' test construction
- Constructing tests according to their own expectations
- Constructing tests according to the students' competence

Themes and Categories

Figure 3. Categories, themes, and theme clusters emerging from the analysis of the interviews with the EFL teachers

6.2. Perceived Purposes of Assessment

The teachers perceived multiple purposes. Specifically, they perceived that the tests were opportunities for them to familiarise their students with the format, content, and procedures of the national examinations. In addition, they believed that assessment informed the teachers and students about their teaching and learning for modification and improvement. They also thought that assessment made both the teachers and students accountable for their work. The teachers reported that they used tests and marks to control their students' learning.

6.2.1. Preparing the students for the national examinations.

The national examinations shaped the teachers' perceptions of what was important to teach and to learn in their school. The teachers said that assessment helped them prepare their students for the national examinations (i.e., the GEDE and the NUEE), especially for the GEDE. This was important to the teachers because "[t]he target for the students in my school is to pass the GEDE" (TB2). Preparation for the national examinations was central to the schools' activities, so "[a]ssessment and meetings in the school always focus on the GEDE" (TB3). The need to prepare the students for the national examinations was expressed by a teacher in School B.

> No matter how their knowledge is, the students need test-taking skills. They need to be familiar with what they have to do in a national examination room. They need to know the procedures in an examination room. They need to know how to communicate with invigilators. In this school, all the students in the same grade take the same 45-minute tests at the same time. This helps the students be familiar with the national examinations. (TB1)

Indeed, TB1's statement suggested that she thought the students should be familiar with the mechanics of taking the national examinations rather than developing test-taking skills. Specifically, she believed that the students needed to be familiar with the procedures used in national examinations.

Some teachers reported that they familiarised their students with the knowledge and language skills tested in the national examinations as well as the format and procedures of these examinations. Some teachers argued that the knowledge and skills that were not assessed in the GEDE and the NUEE should not be assessed in schools. "This is out of my authority, but I think in English instruction, as expected by the MOET, writing should not be assessed. In the GEDE and the NUEE, the students are given multiple-choice tests taken in 60 minutes" (TA1). Another teacher in School A held a similar view.

> I think it is not necessary to assess listening and speaking skills. Why? We have to adhere to the GEDE. Listening and speaking skills are not tested in the GEDE. Our assessment should

adhere to the content and skills tested in the GEDE. We need to prepare our students for the GEDE. (TA3)

In line with this belief, a teacher in School B reported that she focused her 15-minute tests on the knowledge and skills that were tested in the national examinations. She stated, "I often test grammar and reading, mainly grammar, in my 15-minute tests, because the GEDE and the NUEE mainly test grammar and reading" (TB1). These teachers' statements suggested they considered the national examinations to be the MOET's explicit expectations of the knowledge and skills on which school teachers and students needed to concentrate. They believed that they needed to prepare the students to demonstrate the knowledge and skills that they assumed the MOET expected and assessment of other knowledge and skills "was not necessary" (TA3).

The teachers also reported that they familiarised their students with the multiple-choice format. "The students will take multiple-choice tests in the GEDE. Therefore, we give them multiple-choice tests so that they are familiar with the format of the GEDE" (TA1).

The analysis here revealed that the teachers perceived that they needed to prepare their students for the national examinations. The consequence of the focus on this preparation was that the teachers thought that they should narrow the knowledge and skills tested and adopt the format of the national examinations for their classroom tests. Specifically, they focused their tests on grammar and reading because these were tested in the GEDE and the NUEE. They thought that it was unnecessary to assess listening, speaking, and writing because these skills were not tested in the national examinations. However, the results of the analysis of the interviews with the EFL teachers also indicated that the teachers perceived that assessment helped them modify their teaching and helped their students improve their learning.

6.2.2. Modifying and improving teaching and learning.

The teachers perceived that assessment helped them evaluate their own teaching in order to modify it. They also thought that assessment informed them about their students' learning. In addition, the teachers said that assessment helped the students improve their learning through the students' errors and the teachers' feedback on these errors.

First, the teachers thought that assessment helped them evaluate their own teaching. A teacher in School B said, "The second purpose of assessment is for the teachers to evaluate their own teaching. They check whether the students understand their lessons. They check whether their teaching is appropriate to the students' competence" (TB1). This teacher thought that through assessment, the teachers could evaluate whether or not their teaching was appropriate for their students. One teacher in School A explained how assessment informed her about her own teaching.

Through test results, I can see whether my teaching is good or not and whether my students understand my lessons or not. When I am teaching, I ask my students but they often do not raise their hands. Therefore, test results help me see whether my teaching is good or not. If the students understand my lessons, they can do the tests. If they don't, they cannot do the tests. (TA2)

This excerpt suggested that the teacher relied on the students' test results to judge the students' learning and her own teaching.

Second, the teachers thought that assessment helped them modify their teaching. "At the beginning of a lesson, I ask them about the previous lesson to see whether they remember the content of that lesson. Then I decide whether I need to remind them of the content in that lesson or I should start the new lesson" (TB1). In this excerpt, the teacher reported making on the spot decisions about her teaching relying on her judgements about the students' learning. Similarly, a teacher in School A said that assessment informed him about the students' knowledge gaps so that he could help his students fill these gaps. "If we know what the students know and what they do not know, we can supply them with knowledge" (TA1). However, the teachers relied on the students' test marks to make judgements. "Test results are most useful for the teachers because they help them know how their students have learned and they will change their teaching accordingly" (TB3).

Third, the teachers thought that assessment informed them about their students' learning. One of the teachers said, "Assessment is very necessary for teachers. We have to assess the students to know about their competence" (TA3). Similarly, a teacher in School B said, "I think assessment helps me know about the students' competence" (TB2). Some teachers equated "competence" with the students' understanding of the lessons. For example, a teacher in School A said, "[a]ssessment means checking to see whether the students understand what we have taught in order to know their competence" (TA2).

Some of the teachers elaborated on how they assessed their students in order to determine their students' understanding. One teacher in School B said, "After I have taught them something, I ask them questions to check whether they understand it" (TB1). Another teacher in School B stated, "For example, after I have taught my students a grammar structure, instead of asking: 'Do you understand what I have taught you?', I use some questions to check whether they understand my lesson" (TB3). This teacher said that she used multiple methods of assessment but tests and examinations were the primary methods. She stated, "After I have taught a lesson, I use various methods to see whether my students have understood the lesson. However, the most common method is to use tests and examinations" (TB3).

Fourth, the teachers thought that through assessment, they helped the students improve their learning. "After I have taught a lesson, I ask them questions in order to consolidate the content that I

have taught. This consolidation helps the students remember the lesson better" (TB1). This teacher also said that through assessment, she knew the students' errors and she provided them with feedback for learning improvement. She stated, "For example, through the tests, I know the errors that the students often make, then I will help them correct the errors so that they pay more attention to these errors" (TB1). Another teacher in School A also perceived that assessment informed the students about their knowledge gaps and they could learn from their own errors and the teacher's feedback. "Through the errors that they make in the tests and the teachers' feedback, the students know their knowledge gaps and they know what they need to focus on and what they need to review in order to improve their learning" (TA1).

These teachers perceived that assessment informed them about their own teaching so that they could modify it. The teachers also believed that assessment informed them about the students' learning and they helped their students improve learning by providing the students with feedback. However, even though one teacher (i.e., TB3) said that she used various assessment methods, the teachers' reports indicated that they relied mainly on tests rather than on interactions, observations, and other alternative assessments. Their reliance on tests and test results suggested that they assumed that tests were accurate measures of the students' learning. These teachers' accounts also indicated that assessment was used for accountability purposes; this will be discussed in the next sub-section.

6.2.3. Making the teachers and students accountable for their work.

The teachers believed that assessment made themselves and their students accountable for their teaching and learning respectively. A teacher in School A explicitly referred to this purpose. "I think assessment makes us and our students accountable for our work. Without assessment and marks, most students would not try to learn" (TA1). Another teacher in School A argued,

> I cannot imagine what teaching and learning would be if there were no tests and examinations. How could the school administrators, parents, students, and our colleagues judge our teaching without tests and examinations and marks? And how could we make the students accountable for their learning if there were no tests and examinations and marks? Therefore, assessment makes both teachers and students accountable for their work. (TA3)

This argument was supported by a teacher in School B. She stated, "Assessment results make the students and teachers accountable for their learning and teaching because marks are used to make judgements about learning and teaching" (TB2). These statements indicated that the teachers believed that judgements about their teaching and the students' learning were based on the students' marks and thus marks made both the students and teachers accountable for their work. Due to this use of marks, another teacher in School B perceived that "[a]ssessment information is most useful for the school administrators because it is the basis for them to make judgements about the students'

learning and the teachers' teaching" (TB1). However, it may be inappropriate to use the students' marks to hold the teachers accountable for their teaching because the students' marks were influenced by various factors that were beyond the teachers' control (Harlen, 2007).

6.2.4. Summarising learning outcomes for ranking and reporting.

The teachers perceived assessment to have the purpose of summarising learning outcomes for ranking and reporting. A teacher in School B stated, "The ultimate purpose of assessment is to generate the number of marks regulated by the MOET to summarise the students' learning outcomes in a semester" (TB1). A teacher in School A further explained the necessity of summarising the students' learning outcomes. She stated, "How could we know the students' learning outcomes without tests and marks? Right? There have to be tests and marks to summarise the students' learning outcomes for ranking and reporting" (TA2). The statement indicated that TA2 believed that only tests and marks informed the teachers about their students' learning outcomes. Researchers argue that data about students' learning should include both quantitative and qualitative information and various methods of assessment need to be used to obtain such data (J. G. Brooks & M. G. Brooks, 1999; Estaji, 2011). However, this section indicates that the teachers believed that marks were measures of the students' learning outcomes and administering tests in order to obtain marks for ranking the students and reporting to parents was the most important purpose of assessment in their schools.

6.2.5. Controlling the students' learning.

The teachers in these schools reported that they used assessment to control the students' learning. Specifically, they said that they used marks as extrinsic rewards to motivate the students to participate in the class activities and to give the students confidence. In addition, they reported that they used assessment to scare their students in order to make them learn. They also said that they used difficult tests to prevent the students' complacency.

The teachers reported that they used marks to motivate their students to participate in the classroom activities. A teacher in School B thought that her students were motivated by marks and she used marks as extrinsic rewards to engage them in the classroom activities. "The second purpose of assessment is to motivate the students to learn. The students are very excited when I ask them to participate in the class activities and give them marks" (TB3). In a similar vein, a teacher in School A stated,

> In my teaching, I often motivate my students. I use marks to motivate them. In general, the students participate in the class activities for marks. For example, I say, "I will give 10 marks. Who can answer this question?" I know that this way of engaging the students'

participation is for the sake of marks, but it is a way of making the students interested in the
lessons. (TA2)

The above statements suggested that the teacher used marks to engage her students in the classroom
activities rather than using assessment to gather information about their learning.

One teacher reported that she assigned marks to the students according to their participation
rather than according to the quality of their work because she rewarded them for participation in the
classroom activities. "I assign marks according to individual students. For example, some students
deserve 7s, but sometimes I give them 9s when they participate in the classroom activities. I need to
motivate them. In general, I need to be flexible" (TA2). This teacher used marks to involve the
students in her lessons rather than to certify their knowledge and skills. As a consequence of such
practices, marks were not accurate measures of the students' learning outcomes. In addition, some
students were more advantaged than others because the teacher gave marks according to individual
students' participation in her lessons.

Several teachers reported using marks as punishment. A teacher in School A cited two
events in which she used marks to punish the students who did not do their homework.

> For example, last Friday I taught a grammar structure, and I told them to do some research
> into the pronunciations of S and ES endings.
> This morning I checked two students, and they did not have anything in their
> notebooks. I said, "If you can show me your notes about the pronunciations of S and ES
> endings, I will give you 10 marks. Bring me your notebooks." None of the students in the
> class had the notes. I gave these two students a zero each.
> Recently I told the 12th graders, "At home, write a paragraph and prepare an oral
> presentation of the paragraph." After that I checked two students, but neither of them had
> written anything. I told my students, "Raise your hands if you wrote the paragraph." Of the
> 44 students in the class, 38 of them did not write anything; only six of them wrote the
> paragraph. I gave the six students 10 marks each, and I subtracted two marks from the
> previous oral tests of those who did not write anything. (TA2)

This extract indicated that TA2 used marks to reward the students who did their homework and to
punish the students who did not. Therefore, marks were not given as signs of the quality of their
work. When marks were assigned this way, they may have ensured that the students participated in
the class activities, but the practice was unfair when it was used to make judgements about the
students' learning outcomes and to rank the students. In other words, the teachers used assessment
to achieve one purpose (i.e., make students participate in the classroom activities and complete their
tasks) but they sacrificed many other purposes.

The teachers also reported that they used marks to enhance their students' confidence. A
teacher in School B admitted, "In my school, marks are not accurate measures of the students'
learning outcomes. Some students are always absent from class, but they are given 5s in order to
give them confidence" (TB3). Marks were not accurate measures of the students' learning outcomes

because "[m]y marks are more of building up the students' confidence than of measures of their knowledge" (TB3). Not only the teachers in School B, where less competent 9th graders in the city were enrolled, but the teachers in School A, where the most competent 9th graders were enrolled, used marks to build their students' confidence. A teacher in School A explained,

> From many years' experience, my students rarely obtain fail marks in English. I know my students' competence and construct tests appropriate to their competence so that they like English. Even though they are weak, they do not feel disappointed. Who will we teach when our students are both weak and disappointed? We have to make our students confident so that they are not scared when they take tests. (TA3)

TA3 argued that giving students tests that did not fail them was the way she followed to make her students confident and to encourage them to like English. This suggested that she used marks for affective purposes.

The teachers reported using assessment and marks to scare their students. "I also use marks to scare my students. Because marks influence the students' retention and promotion, I use marks to scare them, especially those whose marks are low" (TB3). The teachers' use of marks and assessment to scare students was intended to make them learn. A teacher in School A cited her former teacher as successful in using assessment to scare students in order to make them review their lessons.

> When I was in high school, one of my teachers did not give oral tests in every lesson. Instead, he spent one period (45 minutes) checking what we had learned in two or three chapters. He gave written questions. He asked oral questions. We were really scared, so we learned everything he had taught. (TA2)

TA2 appeared to support assessment that was administered after a period of learning in order to check the students' memory of what they had learned. She believed that students reviewed their lessons because they were scared of being assessed unexpectedly. She said, "I often spend one period asking the students about what they have learned in the last five periods. This makes them scared. They do not know what I will ask them, so they have to review their lessons" (TA2). While researchers have advised that assessment should be embedded in instruction in order to monitor and support learning (Elwood & Klenowski, 2002), for this teacher, assessment administered after a long period of learning made her students review their lessons. In addition, researchers have said that feedback from peers and teachers, self-monitoring, and self-regulation improve student learning (Black & Wiliam, 1998a; Elwood & Klenowski, 2002; Stobart, 2008). However, this teacher believed that pressuring and scaring her students made them try to review their lessons.

Overall, the teachers held different perceptions of the purposes of assessment that were in conflict. On the one hand, they assumed that marks were accurate measures of the students' learning outcomes and believed that assessment informed themselves and the students about their

teaching and learning. They believed that assessment helped the students modify and improve their learning. They perceived that assessment and marks made both the teachers and students accountable. They also thought that tests needed to be administered to obtain marks for ranking the students and reporting to parents.

On the other hand, they used assessment and marks as rewards and punishment rather than measures of learning outcomes. While they said they used marks to build the students' confidence, they also said that they used assessment and marks to scare their students, believing that the pressure created by assessment and marks made their students learn. The next section elaborates the teachers' perceptions of the meaning of assessment and of the stakeholders' expectations and concerns about assessment.

6.3. Marks as the Main Concern

The FEL teachers perceived assessment as a technology used to measure the students' knowledge and to generate marks. Their perceptions of the MOET's, the DOET's, the school administrators', parents', and students' expectations and concerns about assessment indicated that the teachers thought that these stakeholders were mainly concerned about marks.

6.3.1. Assessment as a technology used to measure knowledge and obtain marks.

The teachers regarded assessment as a technology that could be used to measure the students' learning outcomes and to assign marks. According to the teachers' perceptions, assessment meant measuring the students' knowledge. "When thinking about assessment, I think about assessing how much information or knowledge the students have learned after a lesson or some lessons" (TA1). This teacher equated assessment with measuring students' knowledge. A teacher in School B considered tests and examinations to be the primary methods for measuring the students' knowledge. She stated, "When I hear the word assessment, I first think about tests and examinations administered to see how much the students have learned" (TB3). In turn, these results generated by assessment were used to make judgements about students' learning outcomes. "The purpose of assessment is to measure students' knowledge in order to make judgements about their learning outcomes" (TA3).

The teachers thought that assessment meant obtaining marks. A teacher in School B explicitly said that the teachers had to obtain marks to write in the class register. "We administer tests in order to generate marks to write in the class register" (TB3). This perception may result from the requirement of generating the regulated number of marks for each student. A teacher in

School A stated, "In one semester, there are two 15-minute tests, two 45-minute tests, two oral tests, and one end-of-term test. There are seven individual marks" (TA1).

The requirement to generate a set of seven marks for each student pushed the teachers to find opportunities to assign marks. One teacher said, "In principle, a task is part of a lesson, but I have to give them marks when they participate in the task. I have to use many ways so that I have the required number of marks" (TA2). This teacher also said that she had to violate the regulations about assessment in order to obtain the required number of marks for her students.

> According to the regulation, I have to give the students two oral tests and two 15-minute tests. Each week, there are three periods of English. In each period, I can test only one student orally. Each semester consists of around 15 weeks, so I can only give one mark for oral work to each student. I have to do many things to have the required number of oral marks for each student. For example, I give my students 10-minute written tests. This is a violation of the regulation, but I have to do so in order to record two oral marks for each student. (TA2)

The teachers' concerns about administering the mandatory tests to generate the required number of marks was a managerial concern and may subordinate the concern about students' learning improvement.

This sub-section indicates that the teachers considered that assessment was a means by which they could measure students' knowledge and assign marks as measures of students' learning outcomes. The section also indicates that assessment in these schools was mainly used to generate marks for the class register. The need to generate the required number of marks for each student meant that marks became the main concern of the teachers' assessment administration.

Not only were the teachers concerned about marks, they also believed that other stakeholders were mainly concerned about marks. The teachers' perceptions of other stakeholders' expectations and concerns are discussed in the following sub-section.

6.3.2. The teachers' perceptions of the stakeholders' expectations and concerns.

The teachers believed that the MOET and the school administrators expected that marks were accurate measures of students' learning outcomes. They thought that parents and the students expected high marks. The following sub-sections elaborate these perceptions.

6.3.2.1. The teachers' perceptions of the MOET's expectations.

The teachers thought that the MOET expected them to obtain marks that accurately reflected the students' learning outcomes. "I think that the MOET expects the teachers to obtain accurate results of the students' learning" (TA1). Similarly, another teacher in School A stated, "I think that the MOET expects real test results and it does not care about high marks or low marks" (TA3).

Another teacher in School B believed that the MOET expected them to modify their teaching according to the information gathered from assessment. "The MOET expects that through assessment, teachers know their students' competence in learning and they will apply the methods of teaching that are appropriate to their students" (TB2).

However, one teacher in School B thought that the MOET over-emphasised test results and this over-emphasis had negative consequences.

> In my opinion, the MOET is on the wrong track because it over-emphasises test results. The MOET believes that test results reflect the quality of teaching and learning. It believes that a school where most of the students can move on to the next grade is a good school. This assumption results in mark inflation. There are students who are in the wrong class, so the MOET should not over-emphasise test results. Learning outcomes should be measured in the NUEE. At present, mark inflation is pervasive. (TB3)

This teacher believed that the MOET's reliance on test results to make judgements about the quality of teaching and learning resulted in mark inflation. She believed that this way of judging pushed teachers to give marks that were not accurate measures of students' learning outcomes and students moved on to the next grade without being qualified to do so.

These teachers believed that the MOET's expectation that marks were accurate measures of learning outcomes and its expectation that students should move on to the next grade were in tension because when two such expectations existed, teachers artificially raised their students' marks in order to raise the rate of students who moved to the next grade.

6.3.2.2. The teachers' perceptions of the principals' expectations.

The teachers believed that the principals expected them to give the students marks that accurately reflected the students' learning outcomes. For example, a teacher in School A said, "The principal expects that the assessment results accurately reflect the students' competence. This is his greatest expectation" (TA1). Similarly, another teacher in School A stated, "The principal expects the teachers to obtain accurate information about the students' competence" (TA3). A teacher in School B explicitly used the word "marks". She said, "The principal expects that the marks are accurate measures of the students' learning outcomes" (TB1).

The teachers also believed that their principals expected the students to obtain high marks. A teacher in School A said, "The principal expects the students to obtain high marks in assessment" (TA2). A teacher in School B said, "The principal expects that the teachers should not be rigid in assigning marks. He expects the students to obtain high marks from the tests" (TB3). The principals' expectations of the accuracy of marks and their expectations related to the teachers' flexibility to assign high marks indicated a tension because when the teachers thought that they should not be rigid in assigning marks, they may have assigned marks that were not accurate

measures of their students' learning outcomes. That is, the teachers assigned high marks at the expense of the accuracy of the marks as measures of learning outcomes. This was very likely because the students' marks were used to make judgements about their teachers' teaching, to rank the students, and to report to parents and the school administrators.

6.3.2.3. The teachers' perceptions of parents' concerns about marks.

Several teachers believed that parents wanted to know about their children's learning. A teacher in School B stated, "Some parents want to know about their children's competence" (TB2). Another teacher in School B said, "Of course, some parents expect to see accurate information about their children's learning, real marks. They want to know which subjects their children are still weak in so that they can remind their children to make a greater effort" (TB1). These teachers believed that some parents expected that their children's marks were accurate measures of learning outcomes because the marks informed them about their children's learning.

However, some teachers believed that most parents focused their children on obtaining high marks. "Parents do not know how their children learn, but they care about marks" (TA1). Another teacher in School B stated, "Most parents have the same expectation that their children have. They want their children to obtain high marks, as high as possible" (TB1).

A number of teachers thought that parents cared a great deal about marks because marks were high-stakes for their children. "Parents are concerned about their children's marks in a semester and in a school year because marks are used to decide whether their children can move on to the next grade" (TB2). Similarly, another teacher in School B said, "When they see their children's report card and their children's marks are below average, they are very sad. They expect their children to pass tests in order to move on to the next grade" (TB3).

A few teachers thought that parents were concerned about their children's marks because the marks had consequences for the parents themselves. A teacher in School A gave an example to illustrate how children's marks influenced parents' social status. He stated,

> I think parents care about marks. I think they care about marks because, for many parents, their children learn not only for their own future but also for their parents, especially for those parents who have some social position. In any organisation, at the end of a school year, there are certificates of merit and money rewards for the staff's children who are ranked as good or outstanding students in schools. If the junior officers' children are rewarded but the senior officers' children are not, the senior officers are not happy. (TA1)

This teacher believed that children's marks were important to parents because they had a social impact on parents.

The teachers believed that parents' high expectations pushed the children to cheat. One teacher in School A said, "Some parents, though their children's learning is very bad, expect their

children to be good or outstanding students. Their expectation creates pressure on the students, and they try by any means to get good marks to be ranked as good or outstanding students" (TA1). This statement suggested that the students' cheating may be influenced by the wish to obtain high marks to satisfy their parents' expectation.

The teachers' perceptions showed the parents' conflicting concerns and expectations. On the one hand, parents expected that their children's marks were accurate measures of their learning outcomes. On the other hand, they expected them to obtain high marks because the children's marks had consequences not only for the children but also for the parents' social position. The teachers believed that parents' expectations of high marks push their children to cheat in order to obtain high marks.

6.3.2.4. The teachers' perceptions of the students' concerns about marks.

The teachers believed that their students were mainly concerned about marks. A teacher in School A said, "For the students, marks are most important" (TA2). She described how her students bargained for marks. She said, "When I ask a question, my students often ask, 'How many marks, Teacher?' They are very excited" (TA2). This teacher believed that the students were more interested in marks than in participating in the assessment itself.

The teachers reported that obtaining high marks was of great concern to the students. One teacher in School B believed that when the students' took tests, their greatest expectation was to obtain high marks. She said, "Of course, the students expect high marks, as high as possible. This is their greatest expectation" (TB1). Another teacher in School B stated, "What do my students expect from assessment? I think they are only interested in marks. In general, they think they have to learn, they have to be able to answer test questions, in order to obtain high marks" (TB2). Because the students expected to obtain high marks, they had different reactions when they received their marks. For example, a teacher in School A said, "In my opinion, in assessment, the majority of students just care about marks. When they obtain high marks, they shout and clap their hands. When they obtain low marks, they look really sad" (TA1). The analysis in this section suggested that the teachers believed that their students were mainly concerned about marks when they took the tests and participated in assessment activities.

6.3.2.5. The teachers' expectations of their students' marks.

The teachers themselves expected their students to obtain high marks and, at the same time, they expected that marks were accurate measures of the students' learning outcomes. "I want my students to obtain high marks from the tests, but their marks need to be accurate measures of their learning outcomes" (TB2). Similarly, a teacher in School A stated, "Honestly, I expect my students

to obtain high marks from the tests, especially from the 45-minute tests and end-of-term tests. These periodic tests are drawn from the school's test bank, so nobody thinks that I am too generous in assigning marks" (TA2). This teacher thought that the teachers in her school were not trusted and high marks obtained from the tests that the teachers constructed themselves would create suspicions about fairness and accuracy.

The teachers provided several reasons for their expectations of their students' high marks. A teacher in School B stated that she expected her students to obtain high marks because her students' marks had consequences for the students and for herself. "You see, my students' marks are used to rank them and to report to parents, and my teaching is judged according to my students' marks. Therefore, I expect my students to obtain high marks" (TB1). Another teacher in School A said,

> Everyone expects the students to obtain high marks – the principal, parents, the students, and of course the teachers. Why? Because everyone looks at the students' marks to say whether they learn well and are hard-working or not. Everyone looks at the students' marks to make judgements about the teachers' teaching. (TA3)

The teachers' expectations that marks should be accurate measures of the students' learning outcomes (discussed earlier) and the perceived consequences of marks for their students and themselves may be in tension. There may be situations where they assigned a mark that represented their students' learning outcomes but they felt that they should give a higher mark for the students' and their own benefit. One teacher expressed her concern about the tension between these expectations as follows.

> My expectations are contradictory. On the one hand, I expect them to obtain high marks because their marks influence their ranking, even their promotion or retention, and their marks also influence how other people judge my teaching. On the other hand, I expect my students to obtain high marks by themselves, not by cheating. But you see, cheating has a great impact on marks. (TB3)

This teacher thought that obtaining high marks and providing accurate marks were sometimes not possible because some students cheated in order to obtain high marks. In other words, the expectations of high marks may result in cheating, which, in turn, influenced the marks as accurate measures of the students' learning outcomes.

This section indicates that the teachers in these schools regarded assessment as a technology used to measure the students' learning outcomes for the purpose of obtaining marks. The teachers believed that all the stakeholders, except the students, expected marks to be accurate measures of the students' learning outcomes. The teachers also thought that all the stakeholders expected the students to obtain high marks because marks were used to rank the students and to make decisions about the students' promotion and retention. In addition, the teachers believed that the students'

marks influenced their parents' perceptions of their social status and judgements about the teachers' teaching.

The expectation of high marks and the expectation that marks represented the students' learning outcomes created tensions because the former expectation at times resulted in cheating, which influenced the accuracy of the students' marks. In addition, the teachers may give their students high marks because marks had consequences for their students and for themselves. Because marks were high-stakes to both the students and the teachers, the teachers' test administration was controlled and compliance was important. The findings related to these issues are discussed in the next section.

6.4. Control and Compliance

The analysis of the teachers' interviews indicated that they perceived that the school administrators controlled their test administration. In addition, the teachers accepted that they had to comply with the regulations about assessment.

6.4.1. The school administrators' control of the teachers' administration of assessment.

The teachers perceived that the school administrators controlled their administration of assessment. Specifically, they reported that the school administrators controlled the format, the timing, construction, administration of assessment, and marking.

The teachers said that they had to use the required test format. A teacher in School A said, "The teachers have to administer 15-minute tests with 100% multiple-choice items. We do not have any other choices" (TA1). This teacher believed that he did not have autonomy in deciding the test format. This teacher also said that the teachers in his school had to follow the regulations about the number of test items in each test. He stated, "15-minute tests follow the school administrators' regulation. For 15-minute tests, the students are given 20 multiple-choice questions. According to the school administrators' regulation, in 45-minute tests, the students are given 40 multiple-choice questions in each test" (TA1). These statements indicated this teacher perceived that the teachers had to follow the regulations about assessment.

The teachers also reported that they had limited autonomy in deciding the timing of tests. "The timing for the 15-minute and oral tests are flexible, but the 45-minute tests have to be administered at set time. For example, in the 70th period, a 45-minute test has to be given. No matter how you teach, you have to give a test in this period" (TA3). In School B, the same regulation was applied. A teacher in School B stated, "45-minute tests have to be administered at set

time. There is a regulation that there has to be a 45-minute test in that period, and I have to administer a 45-minute test in that period" (TB1).

The teachers said that the principals also controlled their test construction. A teacher in School A cited her own case to illustrate how the principal controlled the teachers' test construction. She said,

> For instance, when I submitted a 45-minute test that I had been assigned to construct, the principal said, "Now, take a seat. In 45 minutes, please answer the questions in the test and write down the answers. See how much of the test you can answer and write down the answers and consider how much your students can do in 45 minutes." I have been in such a situation. (TA2)

This account indicated that the principal controlled the teachers' test construction. In TA2's case, principal A may have thought that she had constructed a test that was too easy or too difficult for the students.

Some teachers thought that the principals controlled test administration. In the following extract, a teacher pointed out how her principal controlled the teachers' administration of test.

> The principal is a very influential person. He controls all the stages in assessment, and he provides the specific directions for the teachers' administration of assessment. For instance, he says that the teachers do not have to ask their students to come to the platform to answer questions about the previous lessons. He tells the teachers to create interesting activities and encourage the students to answer questions and give them marks. (TB3)

The extract indicated that the principal controlled the teachers' implementation of assessment by telling them what they needed to do in classroom assessment. His control suggested that he did not trust the teachers' creation and implementation of classroom assessment. One teacher thought that the principal controlled the teachers' test implementation because he wanted to make sure that the teachers administered the required test types. A teacher in School A said,

> They monitor how the teachers administer tests. Some teachers do not give their students continuous tests. This is wrong, but I think it may happen. When I was a school student, this happened. For quite a long time my teacher did not administer oral tests. Then he collected our notebooks and assigned marks for the oral tests. I think this is officially unacceptable. If the school administrators know, they will complain. Through the teachers' test administration, the principal can see whether the teachers are fulfilling their responsibilities. (TA2)

This teacher thought that teachers were expected to administer the required types of tests, and the school administrators monitored and controlled the teachers in order to make sure that they were administered.

The teachers believed that the school administrators also monitored and controlled their marking. For example, a teacher in School B said,

> The principal expects that 90% of the students obtain passing marks in English at the end of the school-year, although the students' competence is not very good. We have to design tests according to their competence, and show them how to take the tests effectively. (TB3)

The teacher suggested that the principal indirectly controlled the teachers' marking by setting a target for the teachers to achieve, and they had to manipulate their assessment in order to meet this target (i.e., constructing the tests according to the students' competence and showing them how to answer test questions). The principal's target of the rate of passes suggested that he had expectations that were in conflict. On the one hand, he expected the teachers to assign marks that were accurate measures of the students' learning outcomes and he required the teachers to follow the procedures for assessment administration. On the other hand, he expected the teachers to be flexible in assigning marks and even set a passing rate for the teachers to achieve. In such a situation, the teachers may give high marks rather than assigning marks that were accurate measures of learning outcomes.

The teachers thought that the Head of the Section (i.e., the Head of the group which included all the EFL teachers in each school) also controlled the teachers' test administration in general and in marking in particular. A teacher in School A said, "The Head of the Section does not care much about how the teachers administer the oral tests and the 15-minute tests. However, he cares about the marks obtained from these tests. If they are too low, he will question whether the teachers are too strict in marking" (TA2). TA2 suggested that the Head of the Section attributed students' low marks to the teachers' marking. She also believed that the expectation of the Head of the English Section about the students' marks influenced the teachers' marking. She said,

> Recently the Head of the Section complained to a teacher in our Section. After the first semester, the Head of the Section asked that teacher why the marks obtained by her students in different classes were different. He said that the end-of-term test was the same for all the students in the same grade in the school, so the difference may be the result from the teachers' teaching, the continuous tests, and the teachers' marking. It means that the marking has been too strict. Being too strict when marking is not good. (TA2)

This teacher believed that the Head of the Section assumed that the students obtained different marks because of the teachers' teaching, their oral tests, 15-minute tests, and the marking. This assumption was questionable when the students in the classes had different abilities.

TA2 gave her own account to illustrate how the assumptions and expectations of the Head of the English Section had a controlling effect on her assessment. She stated,

> In semester 1, my students' marks were quite low compared with the marks obtained by the other teachers' students. The Head of the Section told me, "You should consider how you assessed your students and see whether you were too strict in marking." He did not say it explicitly, but he seemed to imply that "Your students obtained low marks because you gave them difficult tests and you were too strict in marking." Such an implication influences my assessment. (TA2)

TA2's interpretation of the words of the Head of the Section suggested that she believed that the Head of the Section indirectly controlled her assessment in general and her marking in particular. With this belief in mind, she may have given her students high marks in order to be free of criticism and to avoid unfavourable remarks being made about her teaching and assessment.

The analysis in this sub-section reveals that the teachers believed that the school administrators controlled the teachers' test administration. Specifically, they controlled the format, timing, construction, administration, and marking of the tests. Consequently, the teachers perceived that they had limited autonomy in making decisions about how they administered assessment. The school administrators expected that the marks obtained by different students were not too different. They assumed that the difference in the students' marks was a result of the teachers' teaching, the oral and 15-minute tests, and the marking. As a consequence of this assumption, the teachers constructed 15-minute tests according to their students' competence and this assumption influenced their marking. The teachers could resist or comply with the regulations. However, the teachers' accounts indicated that they complied with the regulations. They also explained the reasons for their compliance. The next sub-section discusses the teachers' compliance and their justifications for it.

6.4.2. The teachers' acceptance of compliance and their justifications.

The teachers reported that they complied with the regulations about assessment that were issued by the MOET, the DOET, and the school administrators. According to a teacher in School A, "assessment in the school cannot be different from what is required in the MOET's and the DOET's documents. Assessment has to be administered according to the regulations" (TA1). TA1 stated, "We have to comply with the regulations about assessment. We cannot do anything different" (TA1). These perceptions of the regulations suggested that the teachers believed that they had very limited autonomy in assessment. Not only did the teachers have to comply with the MOET's and the DOET's regulations, but they also had to follow the regulations imposed by the principal and the Head of the English Section. A teacher in School A said, "All the teachers have to implement the regulations issued by the MOET, by the principal, and by the Head of the Section. We have to comply with these regulations about assessment" (TA3). In School B, the teachers also complied with the school administrators' regulations about assessment. A teacher in School B stated, "I have to administer oral tests, 15-minute tests, and 45-minute tests according to the school administrators' regulations" (TB1).

The teachers reported that they chose to comply with the regulations about assessment rather than resist. They gave several reasons. First, they considered compliance to be good a professional practice. A teacher in School B stated, "It is generally assumed that a good staff member is one who

complies with regulations. Therefore, we should comply with the MOET's, the DOET's, and the school's regulations about assessment. Nobody can complain about me when I fulfil the regulations" (TB3). Compliance with regulations was valued, and the teachers complied with the regulations because they wanted to be considered to be good staff members.

The second reason for the teachers to comply with the regulations about assessment was associated with job security. A teacher in School A stated,

> I must say that I have to choose a safe solution in administering assessment. The teachers have to choose a safe solution. I should not do anything that is different from what is regulated. I am honest. It is best for me to do as other teachers do. I should not go on my own road. It does not matter if what I do (that is different from what is regulated) is good, but if it is not good, I am criticised. (TA1)

TA1 complied with the regulations about assessment rather than administering assessment in his own way because he was afraid of being criticised and complying with the regulations about administering assessment ensured security in his job.

Overall, the teachers perceived that they were controlled by the regulations issued by the MOET, the DOET, and the school administrators. The school administrators controlled the teachers because they did not trust the teachers' professional competence nor their compliance with the regulations. They also controlled the teachers because they were concerned about marks. The regulations and the school administrators' control limited the teachers' autonomy in the administration of assessment. The teachers could not choose the format, timing, and types of assessment. The school administrators' expectation that the students should obtain similar marks pushed the teachers to construct 15-minute tests according to the students' competence. The teachers chose compliance because compliance was valued and compliance ensured job security. Qualities of accuracy, objectivity, and fairness were valued in assessment in these schools because marks were used for multiple purposes (see Competing Purposes of Assessment) and assessment was regarded as a technology used to measure the students' knowledge and all the stakeholders were concerned about marks. However, the teachers thought that these qualities were not attained in their schools. The next section discusses the stakeholders' expectations about accuracy, objectivity, and fairness and the reasons why these qualities were not attained.

6.5. Expectations of Accuracy, Objectivity, and Fairness

The results of the analysis of the interviews with the EFL teachers indicated that accuracy, objectivity, and fairness were valued in assessment in the schools. However, these qualities were not attained because the teachers constructed the tests according to their expectations of the students and according to the students' competence. In addition, the teachers perceived that the students' cheating resulted in the failure to attain these qualities.

6.5.1. Valuing accuracy, objectivity, and fairness.

The teachers valued accuracy, objectivity, and fairness in assessment. For instance, a teacher in School A said, "Accuracy, objectivity, and fairness have to be ensured" (TA1). This teacher emphasised, "The first priority is to be fair to all the students, to be fair in assessment" (TA1). Similarly, a teacher in School B said, "Assessment results should be accurate and objective" (TB3).

Because the teachers valued accuracy, objectivity, and fairness, they valued the measures intended to ensure that these qualities were attained. For example, they supported anonymous test construction. One teacher said, "The teachers do not know who designs the 45-minute and end-of-term tests. It is fair and objective" (TA3). This statement suggested that TA3 believed that anonymous tests helped fairness and objectivity be attained. One teacher considered assessment for teaching and learning to be subordinated to achieving objectivity and fairness. She said, "Each class takes a different 45-minute test, but the tests are chosen randomly from the school's test bank. This is not good for teaching and learning because I cannot assess what I have taught my students. But using the school's tests is good because it ensures objectivity and fairness" (TA2). This teacher knew that using the school's tests deprived her of the opportunities to assess her students on the knowledge that she wanted to assess for the benefit of teaching and learning. However, for objectivity and fairness to be attained, she supported the use of the school's tests. The next sub-section discusses the reasons for not attaining these qualities.

6.5.2. Ethical issues related to the teachers' test construction.

Instead of constructing the tests according to the MOET's standards of skills and knowledge, the teachers reported that they constructed the tests, especially the 15-minute tests, according to their expectations of the students. According to one teacher in School A, "[t]he teachers in this school require more of their students in comparison with the teachers in other schools in the province" (TA1). This statement suggested that in School A, both continuous and periodic tests were constructed according to the teachers' expectations of their students. TA1 further explained why the teachers in his school required more of their students than the teachers in other high schools in the province. He stated,

> Because the students in this school are better than those in other schools, the teachers are quite demanding. Our students' average mark for admission is much higher than that of the students in other schools in the province. That is the reason why the teachers in this school require more of their students. (TA1)

This extract suggested that fairness was not attained in this school. Equal treatment for all students is a requirement of fairness, especially when tests are administered to generate marks for ranking, certification, or selection (Elwood, 2013). However, TA1 believed that the students in School A had

to take the tests which were thought to be more difficult because the teachers expected more of their students compared with the expectations of other teachers in the province of their students.

In School A, the teachers constructed the 15-minute tests with different levels of difficulty due to their different expectations of the students. A teacher in School A said,

> Test difficulty depends on individual teachers. I have observed that my 15-minute tests are more difficult than other teachers' 15-minute tests. . . . As a consequence, my students' marks are a little bit lower than the marks obtained by other teachers' students. Because 15-minute tests are different among the teachers, marks among the classes are also different. (TA2)

TA2's statement indicated that because 15-minute tests constructed by different teachers, there was a lack of fairness. One teacher in School A pointed out the inappropriateness resulting from the teachers' different expectations of their students. "Different teachers design tests with different levels of difficulty because they have different expectations of their students, while the students are judged according to the same standard" (TA1). Students in Vietnam are ranked according to an average of their marks; they are not ranked against each other. In other words, assessment is expected to be criterion-referenced rather than norm-referenced. Therefore, tests need to be constructed according to the MOET's standards rather than according to the teachers' expectations of their students. However, the teachers' reports indicated that students were compared with each other in School A.

Specifically, the teachers in these schools said that they constructed the tests according to their students' competence. "In assessment, we give the students tests that are appropriate to their competence. This means, we have to consider the students' competence in order to construct tests so that the tests are appropriate to them" (TB1). In the same vein, another teacher in School B stated, "The competence of the students in this school is not as good as that of the students in other schools. I think the assessment that we apply is appropriate and it fits the students' competence" (TB2). A teacher in School A made a similar claim. "I design tests according to the students' competence. I teach two Grade 10 classes. The students in Class 10A are better than those in the other class, so I give the students in Class 10A more difficult 15-minute tests" (TA2). These statements suggested that the students with different competence were not treated equally when they took the 15-minute tests, even though the teachers said that they valued fairness in assessment.

Some teachers modified the 15-minute tests in order to advantage their own students.

> Nobody wants their students' marks to be lower than those of other teachers' students. They need to change their tests. For example, after the first 15-minute test, I will change the level of difficulty of the second 15-minute test according to the results of the first test. We should not make our students disadvantaged. (TA2)

This statement suggested that TA2 did not construct her tests according to the MOET's standards. Instead, she set her own standards, and she changed her standards in order not to make her students disadvantaged. The statement also suggested that TA2 was more concerned about her students' marks than the accuracy of marks as measures of her students' learning outcomes compared with the MOET's standards. Another teacher in School A believed that it was appropriate to give tests with different levels of difficulty to students with different competence. He stated,

> In general, the competence of the students in this school and that of the students in Private School X are different, so it is not appropriate to give the students in the two schools tests with the same level of difficulty. That is why the 45-minute tests in this school are more difficult than those in other schools. This is certain. (TA1)

TA1's justification for the teachers' constructing tests according to students' competence was not convincing because students were ranked according to their average mark; they were not ranked against each other.

This sub-section indicates that accuracy, objectivity, and fairness were not attained because the application of the MOET's standards of knowledge and skills was not followed when the teachers constructed the tests. The teachers constructed the 15-minute tests according to their expectations of their own students and according to their students' competence. The teachers were more concerned about their students' marks, so they adjusted the difficulty of their 15-minute tests so that their students were not disadvantaged. Even though they said they valued accuracy, objectivity, and fairness, their tests created a lack of fairness among the students in their schools and across the schools in the province. In addition, the teachers believed that some students cheated, so test results were not accurate measures of students' learning outcomes.

6.6. Discussion

Four major findings can be drawn from the analysis of the teachers' perceptions of assessment in their contexts. First, the teachers perceived that assessment had multiple purposes. Some teachers thought that preparing their students for the national examinations was important. They believed that the students needed to be familiar with the format, content, and procedures of the national examinations. This perception was congruent with one of the principals' perceptions about the purposes of assessment. Preparing the students for the national examinations was important in these schools because, as perceived by the principals, the rate of students passing these examinations had a great impact on the schools' reputation (see Chapter 5). Preparing students for examinations was perceived by teachers in various contexts such as the USA (Kahn, 2000; McMillan, 2001), Turkey (Yildirim, 2004), and Hong Kong (G. T. L. Brown, Kennedy, et al., 2009;

Qian, 2014) as a major purpose of assessment. In these contexts, students' marks obtained from large-scale tests had great consequences for school administrators, teachers, and students.

In addition, similar to the principals and students in thi study, the teachers perceived that assessment informed teachers about their teaching and students' learning and thus helped improve teaching and learning. In other studies, Grade 4 to 7 teachers in south-eastern USA (Davis & Neitzel, 2011), Grade 7 to 12 teachers in Canada (Leighton et al., 2010), primary and secondary schools teachers in the UK (Hargreaves, 2005), primary school teachers in Barcelona (Remesal, 2007), and junior secondary school teachers in Hong Kong (Cheng, 2006) believed that assessment informed teachers and students about their teaching and learning. In many studies conducted in different countries, teachers perceived that assessment helped improve teaching and learning. Teachers in New Zealand (G. T. L. Brown & Harris, 2009), in New Zealand and Queensland (G. T. L. Brown & Lake, 2006), in the Netherlands (Segers & Tillema, 2011), and in Hong Kong (G. T. L. Brown, Kennedy, et al., 2009; Cheng, 2006) believed that assessment helped improve teaching and learning. The finding in this study and in previous studies carried out in many countries indicated that teachers perceived that assessment had formative purposes, no matter whether assessment in their contexts prioritised these purposes or not.

The teachers in the present study believed that assessment made teachers and students accountable. This perception was also held by the principals and students in this study. In other contexts such as Israel, Israeli EFL teachers in elementary, junior-high, and high schools also believed that both national standardised tests and alternative assessments were a tool for monitoring and surveillance (Inbar-Lourie & Donitsa-Schmidt, 2009). The accountability purpose was also perceived to be a dominant purpose by teachers in other countries such as New Zealand (G. T. L. Brown, 2004; G. T. L. Brown & Harris, 2009), Hong Kong (G. T. L. Brown, Kennedy, et al., 2009), the Netherlands (Segers & Tillema, 2011), and south-eastern USA (Davis & Neitzel, 2011). This purpose was held by teachers in contexts where assessment had low stakes (e.g., New Zealand (G. T. L. Brown, 2004; G. T. L. Brown & Harris, 2009)) and in contexts where assessment had high stakes such as Vietnam (this study) and Hong Kong (G. T. L. Brown, Kennedy, et al., 2009).

The teachers in the present study perceived that assessment was used to summarise learning outcomes in order to rank students and report to parents. This purpose was congruent with policy-makers' regulation. It was also similar to the purpose perceived by the principals and students in this study. The purpose of reporting to parents was also held by Grade 4 to 6 teachers in South Africa (Pryor & Lubisi, 2002) and Grade 4 to 7 teachers in south-eastern USA (Davis & Neitzel, 2011). Secondary school teachers in Hong Kong (Cheng, 2006) expected that by reporting their students' learning outcomes to parents, they would get support in students' learning from parents.

The teachers in the present study believed that by using marks as rewards and punishment they could control their students' learning. This purpose was associated with the teachers' perception that marks had serious consequences for their students. This purpose was also perceived by the students, who believed that their teachers used assessment and marks to monitor and control them in learning. Using grades to reward and punish students for their effort was also found among teachers of Grade 4 to 7 students in south-eastern USA (Davis & Neitzel, 2011).

The second major finding in this chapter was that the teachers were concerned about marks. A number of the students in this study believed that their teachers cared a lot about marks because students' marks influenced how the teachers' teaching was judged. Some teachers equated assessment with using tests and examinations to measure students' learning outcomes and assign marks. This perception of assessment indicated that these teachers put more weight on assessment of learning than assessment for improving teaching and learning. Junior secondary school teachers in Hong Kong (Cheng, 2006) also equated assessment with tests and examinations. Therefore, they believed that alternative assessments should not be the major assessments in their curriculum. The teachers in the present study expected their students to obtain high marks from tests, and some believed that their principals, students, and parents were also concerned about marks. These findings were new in research studies that have examined teachers' perceptions of assessment.

The third finding reported in this chapter was that the teachers perceived that they were controlled in implementing assessment and they did not have autonomy to decide the format and timing of the tests, especially 45-minute written tests and end-of-term written tests. Additionally, they were not allowed to construct periodic tests for their own classes and were required to anonymously mark test papers from periodic tests. The teachers reported that they complied with the regulations and procedures of assessment issued by the MOET, the DOET, and the school administrators. This meant that their assessment was controlled by administrators at every level. Indeed, the principals reported that they controlled their teachers' assessment and expected their teachers to comply with the regulations about assessment because they wanted accuracy, objectivity, and fairness to be achieved and they distrusted their teachers professional competence, professional commitment, and fairness. Similar findings were obtained from a study about assessment implemented by teachers in social studies in 81 high schools in 33 provinces in Turkey (Yildirim, 2004). Teachers in Turkey were not encouraged to construct multiple-choice tests because their school administrators did not trust their expertise in designing and using multiple-choice tests. These teachers were also not encouraged to use essay tests. They mainly used short-answer tests and occasionally used oral tests (Yildirim, 2004). The teachers in the present study

chose to follow the regulations and procedures for assessment because they thought that compliance was valued and it ensured job security.

The last major finding was that there were contradictions between the teachers' perceptions of objectivity and fairness in assessment, their prevention of the students' cheating in the observed classrooms, and their reported test construction. On the one hand, they valued objectivity and fairness in assessment and expressed support for measures proposed to ensure that objectivity and fairness were achieved in assessment. On the other hand, they reported constructing continuous tests according to their own expectations of their students and according to their students' competence and modified the continuous tests in order to raise their students' marks. The teachers equated fairness with equal treatment of all students in testing. This perception of fairness was consistent with a perception of fairness held by English language teachers in secondary schools in Ontario, Canada (Tierney, 2014). However, teachers in Tierney's (2014) study thought that fairness also meant giving students ample opportunity to learn and to demonstrate learning. In addition, they thought that student learning had to be judged holistically. They believed that learning expectations and assessment criteria were shared with students. Therefore, the perception of fairness held by the teachers in the present study reflected a context of assessment where accurate and objective test results rather than opportunities for students to learn were valued.

This chapter has presented the teachers' perceptions of assessment in their contexts. The next chapter discusses the contexts of assessment in the schools as perceived by the students.

Chapter 7. The Context of Assessment as Perceived by the Students

7.1. Introduction

This chapter presents the results of the analysis of the 12 focus group interviews with the students. For convenience reference, each student was assigned with a code. In TA1S1, for example, TA1 indicated the code for a teacher (T stood for Teacher, A indicated that the teacher came from School A, and 1 indicated the number assigned to the teacher). S1 indicated the code for a student (S stood for Student and 1 indicated that the student was assigned with number 1 among 12 participating students of teacher TA1).

The themes and theme clusters emerging from the analysis of the focus group interviews are summarised in Figure 4. The chapter is organised according to Figure 4.

The first cluster discusses the students' perceptions of the purposes of assessment. The second cluster discusses the students' concerns about marks. Specifically, this cluster presents the students' perceptions of the accuracy of marks as measures of learning outcomes, their concerns about marks, reasons for their concerns about marks, and consequences of their concerns about marks. This cluster also discusses the students' perceptions of their teachers' and parents' concerns about marks. The third cluster addresses the students' distrust of the teachers' fairness and their perceptions of the teachers' feedback to their errors.

7.2. The Students' Perceptions of the Purposes of Assessment

The students perceived that assessment had multiple purposes. First, they believed that assessment made them accountable for their learning. Second, the students believed that the teachers used assessment to monitor and control their learning and to obtain marks for ranking. Third, they thought that assessment informed themselves and their teachers about their learning. They considered assessment to be opportunities for them to improve their learning through reviews of their lessons, errors made in the tests, and feedback from the teachers and friends. The students also believed that assessment helped the teachers evaluate their own teaching in order to modify and improve it.

Figure 4. Results of the Analysis of the Focus Group Interviews with the Students

The Context of Assessment as Perceived by the Students

The Students' Perceptions of the Purposes of Assessment	The Students' Perceptions of the Stakeholders' Concerns about Marks	The Students' Perceptions of the Teachers' Fairness and Feedback	Clusters

Monitoring and controlling the students
- Making the students accountable
- Monitoring and controlling the students

Obtaining marks for ranking
- Obtaining the mandatory number of marks
- Obtaining marks for ranking

Informing learning and teaching for modification and improvement
- Helping the students evaluate their own learning
- Learning from peers' and the teachers' feedback
- Informing the teachers about the students' learning
- Helping the teachers evaluate their own teaching for modification and improvement of teaching

The students' concerns about marks
- The students' perceptions of the accuracy of marks
- The students' concerns about marks
- Demonstrating competence subordinated to obtaining high marks
- Reasons for the students' concerns about marks
- Test anxiety as a consequence of the concern about marks

The students' perceptions of the teachers' and parents' concerns about marks
- The students' perceptions of the teachers' concerns about marks
- The students' perceptions of parents' concerns about marks

The students' distrust of the teachers' fairness
- Distrust of the teachers' fairness in test construction
- Distrust of the teachers' fairness in marking

The students' perceptions of the teachers' feedback
- Written feedback: ticks, crosses, and marks
- The students' dissatisfaction with the teachers' feedback

Themes and Categories

Figure 4. Categories, themes, and theme clusters emerging from the analysis of the 12 focus group interviews with the students

7.2.1. Monitoring and controlling the students.

Some students believed that assessment was used to make them accountable for their learning. One student said, "Assessment makes us learn" (TB1S9). The following conversation suggested that some students thought that assessment made them accountable for their learning.

TA1S3:   Would you learn if there were no tests?
TA1S4:   No. There must be tests.
TA1S6:   But I love learning without tests.
TA1S4:   No. Sorry. Nobody would learn.
TA1S3:   = No =
TA1S2:   = TA1S6 is interested in English =
TA1S6:   = I mean I want to learn without tests, but during the learning process, the teachers assign marks. I think it is not necessary to give tests. Tests create pressure for students. I am telling the truth.
TA1S4:   I think it depends. Without tests, only those students who are keen on English would learn it. The classmates who sit next to me would not learn English if there were no tests. They learn Mathematics, Physics, and Chemistry, but they would not learn English, right?
TA1S3:   Normally I =
TA1S4:   = normally she (TA1S3) does just the same =
TA1S3:   = I will take English in the NUEE, so I spend a lot of time learning English.

Most of the students participating in the above excerpt believed that assessment made the students learn. TA1S3's and TA1S4's statements suggested that tests, especially the national examination (i.e., the NUEE), made the students learn. TA1S6 suggested that assessment should be embedded in instruction. However, this student equated assessment with assigning marks. Concerning the accountability purpose of assessment, some students thought that assessment made the students review their lessons.

TA2S1:   Normally, students do not review their previous lessons, but they have to review their lessons for tests.
TA2S2:   He is telling the truth.
Interviewer: You mean assessment makes students =
TA2S1:   = learn =
Interviewer: = learn regularly =
TA2S1:   = review their knowledge.

This extract suggested that some students believed that tests pushed them to review their lessons.

Several students thought that their teachers used assessment as a means of monitoring and controlling their learning. "The teachers want to see whether or not the students pay attention to their lessons in the classroom" (TB2S2). In addition, another student believed that the teachers administered assessment to make sure that their students completed set tasks. "First the teachers want to see whether or not the students review the lessons and do their homework" (TB3S3). Another agreed, saying, "The teachers assess their students in order to see whether or not the

students review their lessons and prepare the new lesson before going to class and try to remember their lessons" (TB3S2). Some students even thought that the teachers used assessment to scare the students into behaving in the classroom.

TB2S1:   Some teachers use assessment to scare their students.
TB2S5:   When some student is making noise in the classroom, the teacher asks him to answer her questions so that the student obtains a low mark.

This sub-section suggests that some students believed that assessment made them review their lessons and make an effort to learn. This was especially true for those subjects that they did not take in the NUEE. A number of students thought that the teachers used assessment to make their students review the lessons, do homework, prepare new lessons before going to class, and pay attention to their lessons. In other words, assessment made the students accountable for their learning.

7.2.2. Obtaining marks for ranking.

Most students considered assessments to be a means for the teachers to generate the required number of marks.

Interviewer: Any other purpose for the teacher's assessment?
TA1S4:   To obtain the required number of marks.
Interviewer: To generate =
TA1S3:   = generate enough marks =

Many students said that the teachers administered tests in order to obtain marks to rank them. One student said, "My EFL teacher administers tests in order to generate marks for the score book" (TB1S11). Another student said, "After each semester, students are ranked, so the teachers assess us in order to obtain marks" (TB3S5). Similarly, another student said, "The teachers administer tests in order to obtain marks for ranking the students" (TA2S2). These students perceived that assessment had managerial significance and they perceived that assessment was directly related to their ranking.

7.2.3. Informing learning and teaching for modification and improvement.

Many students thought that assessment informed the students about their learning and helped them improve learning through review, errors, and feedback from the teachers and peers. They also believed that assessment informed the teachers about their teaching and helped them modify their teaching in order to help the students improve their learning.

A number of students believed that assessment helped them evaluate their own learning. "I want to evaluate the skills and knowledge that I have" (TB3S3), or "I want to see how much I have

understood" (TB3S8). These students regarded assessment as a tool that helped them measure their knowledge and skills.

In addition, some students thought that they learned from the errors they made in the tests through their friends' and teachers' help.

TA2S2:   For example, after we have made an error, we learn something for the next test.
Interviewer: But some of you said that your teachers just gave you 5 or 7 marks and no correction at all. What could you learn from such tests?
TA2S3:   We can look at the test papers with good marks.
TA2S2:   We can compare our answers with those of the students who are given good marks.
Interviewer: But learning is not as simple as choosing A, B, C, or D. Learning comes from why B or C is chosen =
TA2S1:   = we can ask our classmates, "Why do you choose that response?"
TA2S2:   = or we can ask the teacher. She will explain.

These students believed that they learned from their errors by comparing their incorrect answers with the classmates' correct answers. Additionally, they looked for help from their classmates and teachers in order to improve their learning. These students believed that feedback was important in learning.

The majority of students believed that the teachers wanted to check their retention of the lesson content. "The teacher assesses us in order to see whether or not we remember what we have learned" (TB1S11). Another student believed that the teachers wanted to check their understanding. "The teachers assess us in order to see whether we understand what they teach or not" (TA2S4). Another student said "The teachers assess us in order to see how much we understand" (TB3S6). These statements indicated that some students believed that their teachers assessed them because they wanted to know about the students' learning.

Several students thought that the teachers' attempts to establish what the students had learned was not an end in itself. They believed that through assessment, the teachers formed judgements about the students' learning and identified those students who needed more help. "The teachers assess their students in order to know the students' competence. Through the tests, they know how many students are good, how many students are not good, and who need help" (TA2S12). Another student thought that the teachers assessed their students in order to judge their learning and to help them improve their learning.

TA1S11:   I think through assessing the students, the teachers know how much the students understand.
Interviewer: Why do you think the teachers want to know how much their students =
TA1S11:   = they know about their students' learning so that they can improve their students' learning.

Many students thought that assessment helped the teachers evaluate their own teaching and modify their teaching for the improvement of learning and teaching. One student stated, "Through assessment, she knows the effectiveness of her teaching and tries to make her teaching better" (TA3S7). In a similar vein, another student stated,

> I think assessing students is a way for the teachers to evaluate how they have taught and transferred knowledge to the students. The teachers can see whether or not the students can follow the way they transfer knowledge so that they can modify their teaching . . . (TA2S2).

These students perceived that assessment was helpful to the teachers in improving their teaching.

A number of students elaborated how the teachers modified their teaching using assessment data. "The teachers administer the tests in order to modify their teaching. If we have mastered the lessons, their teaching will be different; if we do not understand, their teaching will be different" (TA3S2). Another student believed that the teachers modified their teaching according to the students' test results. "For example, if the students obtain low marks, the teachers will modify their teaching and try to find the teaching method that is most appropriate for their students so that they can understand the lessons better and obtain higher marks" (TA3S7). This statement suggested that the student believed that the teachers relied on the students' marks in order to modify their teaching.

This section indicates that the students thought that assessment made them accountable for their learning. In addition, they perceived that assessment was administered to obtain marks for ranking them. They also believed that assessment informed them about their learning and they could improve their learning by learning from the errors they made in the tests and get help with fixing the errors from peers and the teachers. The students thought that assessment helped the teachers evaluate their own teaching and modify their teaching in order to improve student learning. They believed that marks obtained from the tests informed the teachers about the students' learning and the teachers relied on marks to modify their teaching. These purposes may compete with each other because when the tests were administered to obtain marks for ranking the students, both the students and teachers may try to obtain high marks by any means (Stobart, 2008). In addition, when assessment is administered for both accountability and improvement purposes, the accountability purpose may subordinate the purpose of learning improvement because "the higher stakes assessment will generally subvert the lower stakes practice" (Johnston & Costello, 2005, p. 264).

7.3. The Students' Perceptions of the Stakeholders' Concerns about Marks

The students had different views about the accuracy of marks as measures of learning outcomes. However, in general, they were concerned about marks, and they thought that their peers were also concerned about marks. Some students explained the reasons for their concerns about marks and for their preference for multiple-choice tests, even though they believed that multiple-

choice tests were not the best methods to use to obtain accurate measures of learning outcomes. Some students attributed test anxiety to the concern about marks. Many students believed that their teachers and parents were mainly concerned about marks.

7.3.1. The students' concerns about marks.

The students held different perceptions of the accuracy of marks as measures of learning outcomes. However, most of them were concerned about marks because they perceived that marks had consequences for them.

7.3.1.1. The students' perceptions of the accuracy of marks.

The students' perceptions of the accuracy of marks as measures of learning outcomes were different. Some students believed that marks were accurate measures of learning outcomes, while others thought that they were not.

Several students believed that marks reflected their competence in learning. "It depends on the individual students' competence. If I learn well, I obtain good marks. Marks depend on my competence" (TB1S5). Similarly, another student argued that marks were accurate measures of the students' learning outcomes because "good" students and "lazy" students were differentiated by marks. "I think marks are accurate measures of learning outcomes because there are 4s and there are 8s. Good students obtain 8s, and lazy students obtain 4s" (TB2S2). Another student believed that marks indicated how the students learned a subject and whether or not they liked it. "If we learn well, our marks are good. If we are not good at or do not like a subject, we do not obtain good marks. If we learn and we like a subject, our marks in that subject must be good" (TA1S7).

However, many students thought that marks were not accurate measures of learning outcomes and other students expressed doubts about their accuracy. Following are some illustrative excerpts. Some students did not believe in the accuracy of marks.

Interviewer:	Do you think marks reflect your learning outcomes?
TA1S1:	No, no =
TA1S6:	= no.

Other students thought that whether marks were accurate measures of learning outcomes or not depended on the individual student's competence.

Interviewer:	Do you think marks are accurate measures of your and your friends' learning outcomes?
TA1S8:	Perhaps no
TA1S9:	Just partly.
TA1S7:	Partly.
TA1S9:	For good students, marks are accurate measures of their learning outcomes, but for less competent students, they are not.

The students gave various reasons why marks did not accurately reflect students' learning outcomes. One of these reasons was cheating.

Interviewer: Do you think marks were accurate measures of your and your friends' learning outcomes?
Students: No
Interviewer: Why not?
TB1S11:   Because we can ask our friends for answers while we are taking a test.
Interviewer: Any other reason?
TB1S7:   We use materials.
Interviewer: Materials.
TB1S9:   It is easy to cheat in the classroom.
Interviewer: Cheating? What else?
TB1S9:   We give each other answers and materials. Many ways.

In addition to cheating, the students thought that good luck and making random choices affected the accuracy of marks.

TB2S5:   Copying is just one reason. Sometimes I make errors just by bad luck =
Interviewer: = Bad luck and good luck?
TB2S5:   Yes, we tick at random. We can be right by good luck, or we can be wrong by bad luck.

In the following excerpt, some students reported that the pressure to do well in testing affected their performance.

TA3S4:   For some students, marks are accurate measures of their learning outcomes; for other students, marks are not accurate measures of their learning outcomes.
Interviewer: Why?
TA3S4:   I have observed that many students who do not specialise in English obtain higher marks than those students who specialise in English, so I think marks are not reliable measures of learning outcomes.
Interviewer: Do you know why =
TA3S3:   = because there is pressure when we take a test =

Although the students did not share the same beliefs about the accuracy of marks, many students believed that marks were not accurate measures of learning outcomes because some students cheated, they made random choices when they completed a test, and the pressure to do well in testing influenced their test results.

7.3.1.2. The students' concerns about marks.

Several students said that they were not concerned about marks. These students believed that they were more concerned about learning and knowledge. "My knowledge is more important. I care more about the knowledge that I have. I do not care about marks" (TB2S4). Another student stated, "Marks are not important. The knowledge that I show in the tests and the knowledge that I remember are more important than marks because I can copy to obtain good marks" (TB2S2). This student's statement suggested that cheating was within the students' control.

While some students reported that they did not care much about marks, many of them considered marks to be their main concern in assessment.

Interviewer:	When you are assessed, what do you expect?
TB1S6:	I always want to obtain high marks =
TB1S1:	= marks, high marks =

Similar responses were recorded in another focus group.

Interviewer:	When you are assessed, what do you expect?
TA1S6:	High marks.
TA1S1:	High marks.
TA1S4:	First, high marks.
TA1S1:	Marks, as high as possible.

The expectation of high marks was very common among the students when they responded to the question about their thinking when they heard the word "assessment".

TA1S5:	Marks
TA1S3:	Nervous.
TA1S1:	Marks.
Interviewer:	Marks. You think about marks.
TA1S3:	Assessment is not necessarily related to marks.
TA1S6:	But tests are given marks.
TA1S5:	But most students are concerned about marks. I do not know about the students in other places, but the students in my class =
TA1S4:	= my class are concerned about marks =
TA1S5:	= my class are the same, and many other classes are just the same.

Although one of the students in the above extract (i.e., TA1S3) argued that marks were not a necessary part of assessment, many students associated assessment with marks. Specifically, they thought that marks should be given when the students were assessed.

Interviewer:	Your statements suggest a question. Do you think your teacher should give you tests in which you are allowed to use materials and talk to your classmates and no marks are given?
TA1S1:	No.
TA1S4:	Yes, I think there should be this type of tests.
TA1S2:	I think the teacher should not give us such tests because if no marks are given, the students will not care, will not try to do their test =
TA1S4:	= I meant there should be such tests so that we can learn, but many students would not take the tests seriously if marks were not given =
TA1S2:	= when the students are informed that marks are given, even though finally the teacher does not assign the test papers with marks, they take the test seriously. They refer to books, Google =
TA1S4:	= we use our smartphones to search Google.
Interviewer:	You mean the students take tests seriously only when marks are given =
TA1S4:	= yes. When the teacher says he will give marks, the students fill pages with words =
TA1S2:	= fill pages with words =
TA1S3:	= no mark, no word but if marks are given, we write many words.

Apart from one student in the above extract (TA1S4), other students participating in the conversation thought that the teachers should assign marks when they assessed them. They thought that marks engaged them when taking the tests. This suggested that many students were concerned about marks and they did not see these tests as opportunities to demonstrate what they knew and could do or as learning events. This extract suggested that when assessment was associated with marks, the students paid more attention to marks than to the purpose of learning from assessment.

7.3.1.3. Demonstrating competence subordinated to obtaining high marks.

The students' preferred methods of assessment and the methods that they believed would gather reliable information about learning outcomes were different. This indicated that they paid more attention to obtaining high marks than to opportunities for the teachers to gather accurate data about their learning outcomes.

Many students believed that, for the purpose of making judgements about their learning outcomes, the teachers' frequent observations during instruction were better than tests. One student's remarks reflected this view.

> I do not believe in written tests because, as my friends said earlier, they do not accurately measure learning outcomes due to cheating or luck. I think the teachers' observations of the students' learning during the process of instruction is better than tests in terms of judging our learning outcomes. (TA3S3)

Some students believed that peer-observation and social interaction were better tools than tests regarding the purpose of making judgements about students' learning outcomes. They argued that their classmates knew their learning better than the teacher because they had close contact and more frequent observations of each other's learning, while the teacher did not have such opportunities.

> Interviewer: Do you think the teacher or your classmates know your real competence better?
> TB1S7: Classmates.
> TB1S9: Classmates.
> TB1S11: Of course, classmates.
> Interviewer: Why doesn't your teacher know your learning better than your classmates?
> TB1S10: Because she gives us tests in order to judge our real competence but we are not honest. She does not know whether our marks are real or unreal.
> TB1S12: Each week the teacher meets us in only three periods, while we always learn with each other and help each other. The teacher does not know for sure about our learning.

Some other students believed that direct exchanges or interactions between the teacher and the students helped the teacher know about their students' learning. "The teacher should assess the students by using oral questions and answers" (TA2S2). Another student stated, "During the lessons, the teacher asks questions and gives marks in order to judge their competence, if the

students can answer the teacher's questions" (TA3S3). This student mentioned an assessment method that researchers (e.g., Klenowski, 2009) recommend (i.e., dialogues).

However, most students preferred to take multiple-choice tests for various reasons. First, they said it was easier for them to cheat when taking multiple-choice tests. "It is easier for us to copy when taking multiple-choice tests" (TB2S4). Similarly, another student said, "When we take a multiple-choice test, if the questions are too difficult, we can ask other students and they can answer" (TB1S8).

Second, many students said that it was easier to respond to multiple-choice items because the responses were predictable. "We can predict the correct responses when we take multiple-choice tests" (TA1S1) and "the responses in multiple-choice tests are suggestive so we can predict the correct responses" (TB1S8). The students also said that they preferred this type of tests because they could tick the responses randomly and had a chance of selecting correct answers.

TA2S9:   We can tick the responses by random choice.
TA2S10:   It is easier to tick the responses by random choice.
TA2S11:   When taking a multiple-choice test, if we do not know the correct responses, we can tick the responses by random choice when we do not have enough time.

Another student stated, "When we do not have enough time, we can tick the responses by random choice, and we have 25% of having a correct response" (TB1S11).

The third reason was that they did not have to face their limited ability in writing and speaking. "Like my friends, I prefer multiple-choice tests because it is more risky to write full sentences in open-ended tests" (TA3S4). Another student stated, "Multiple-choice tests are best. We do not have to write. When we write, we make grammar errors" (TA1S1). A student shared a similar reason.

We know that it is not easy to tick 100% of the correct answers in multiple-choice tests and obtain 10 marks, but we prefer multiple-choice tests to open-ended tests because many of us are not good at English. We do not have a good vocabulary to write full sentences. (TB1S4)

In explaining why the students preferred multiple-choice tests to oral tests, one student said, "When we answer questions orally, our pronunciation is not good enough, so we are afraid of oral tests" (TB1S8). These students preferred multiple-choice tests because they did not want to obtain low marks on account of their lack of confidence in their writing abilities and their limited command of English.

The analysis in this sub-section indicated that although the students considered the teachers' frequent observations, oral questions and answers, close contact, and direct interactions to be reliable assessment methods for the teachers to know about their learning, they preferred multiple-choice tests for various reasons. The reasons they mentioned suggested that they cared about marks.

7.3.1.4. Reasons for the students' concerns about marks.

The students gave a number of reasons for their concerns about marks. First, some students perceived that marks were very important to them. "I think marks are very important, so I have to have the marks that I want by any means. Marks have to be over 5 because they have consequences for us" (TA1S4). They described their different responses to marks. "I do not care how I obtain the marks, but I am very happy and I tell everyone when I get 8s or 9s, but I am very sad when I obtain low marks" (TA1S3). This statement suggested that this student was concerned about marks, no matter whether they were accurate measures of learning outcomes or not.

One student said that marks were important because they were used to make judgements about their learning.

Interviewer:	What about you? Are marks important to you? If they are, why?
TA3S11:	Yes, they are very important because they influence everything. People look at our marks to make judgements about our learning. They do not care how we learn and whether we obtain the marks by ourselves or by cheating.

This student perceived that learning was judged solely by marks, no matter whether they were obtained by fair means or foul. This explained why many students "do not care how [they] obtain the marks" (TA1S3) and they "have to have the marks that [they] want by any means" (TA1S4).

In addition, many students were concerned about marks because marks influenced their ranking. One student said, "I want to obtain high marks in order to get a high ranking" (TB3S5). Similarly, some other students said,

Interviewer:	When you take an English test, what do you expect?
TA3S5:	Marks.
TA3S6:	Marks.
Interviewer:	Can you tell me why?
TA3S3:	Although marks are not accurate measures of our competence, they influence our ranking in official paperwork.

The students' concern about marks was understandable because marks were used to rank the students and the ranking was written in their official record and was regarded as evidence of their competence in schooling.

Some students believed that their ranking influenced other people's views of them.

Interviewer:	How important do you think your marks are?
TA2S4:	The students think =
TA2S3:	= marks influence our ranking =
TA2S1:	= being ranked as a good student is different from being ranked as an average student.
TA2S3:	In everyone's eye, an outstanding student, a good student, and an average student are different =
TA2S2:	= Not being looked down on or scolded =

These students believed that the students with different rankings received different levels of respect from people, and their image was ruined if their ranking was low. With these beliefs about the influence of marks and ranking on their self-esteem, obtaining high marks by any means was unacceptable but understandable.

Some students thought that marks also influenced their parents' reactions. "I want high marks in order to be ranked as a good student so that my parents will give me a reward (TB2S5). Another student said that her marks influenced her parents' behaviours towards her. "Marks are important because my parents are happy when I obtain high marks, and they scold me when I obtain low marks" (TA2S8).

Moreover, the students perceived marks to be part of competition among each other. "I agree with my friend that high marks make parents happy. What is more, marks are an essential part of the competition among us in learning" (TA2S10). Another student said, "Tests should be assigned with marks so that students compete with each other" (TA2S12). As discussed earlier, it was possible that many students competed with each other to obtain high marks because marks and ranking influenced their self-image. Additionally, marks influenced the students' judgements of each other.

Interviewer:	Do you judge each other according to marks?
TB2S8:	No.
TB2S7:	Why not?
TB2S10:	We do.

Although not all the students competed with each other, competing for high marks was part of their daily learning experiences.

This sub-section indicates that the students held different perceptions of the accuracy of marks as measures of learning outcomes and whether or not the students were concerned about marks. Many students were concerned about marks because marks had various influences on their rankings, their emotions, self-esteem, and treatment from parents. Because of these perceived impacts of marks, some students reported that they were ready to cheat in order to obtain high marks.

7.3.1.5. Test anxiety as a consequence of the concern about marks.

Test anxiety is defined as "the set of phenomenological, physiological, and behavioral responses that accompany concern about possible negative consequences or failure on an exam or similar evaluative situation" (Zeidner, 1998, p. 17). Many students in this study reported that they felt anxious as a consequence of their concerns about test marks. In the following extract, the students described their feelings when they thought about assessment.

Interviewer:	What do you think when you hear the word "assessment"?
TB1S12:	Worried.
Interviewer:	Worried.
TB1S10:	Nervous.
Interviewer:	What else?
TB1S9:	Feeling tired.
Interviewer:	What else? What about you?
TB1S7:	Scared.
Interviewer:	Scared?
TB1S7:	Yes.

Words such as "worried," "nervous," and "scared" suggested that these students felt the anxiety created by assessment. These words also occurred in the descriptions from the students in another group.

Interviewer:	What do you think when you hear the word "assessment"?
TA1S1:	Really scared.
Interviewer:	Scared. What else?
TA1S4:	Feeling worried about the level of the test.
Interviewer:	Worried about the level. What is the level about?
TA1S4:	The level of difficulty of the test.
Interviewer:	What else?
TA1S6:	How difficult or easy the test is.
TA1S4:	Nervous.

The students participating in the above conversations expressed the same feelings (i.e., scared, worried, nervous) because they worried about the difficulty of the tests, which may influence their performance on the tests and their marks.

In the following discussion, the students referred to the pressure that they felt from assessment.

Interviewer:	What else?
TA3S3:	Pressure.
TA3S2:	Marks for evaluating the students' learning outcomes.
Interviewer:	Pressure. Why is there pressure?
TA3S3:	In general =
Interviewer:	= speak up, please =
TA3S3:	= not only English =
Interviewer:	= assessment in general =
TA3S3:	= in general, I feel pressure when I take tests in the subjects that I am not good at =
TA3S4:	= there is pressure in any subject =
Students:	(laugh)
TA3S3:	= right, in general, there is pressure.

The above extract indicated some students felt the pressure created by the tests, especially by the tests in the subjects that they were not good at.

When the students were asked whether the teachers should give them tests for which marks were not assigned, they said they liked this way of assessment. The reason they suggested was that such tests decreased the pressure created by marks.

Interviewer: Do you think your teacher should give you tests without assigning marks?
TB1S9: Yes.
TB1S10: Yes.
TB1S11: If so, the pressure created by marks will be reduced.
TB1S12: In this case, we are more independent because no mark is assigned. We do not know what our test results will be, but we are independent.

By "independent", TB1S12 suggested that the students did not rely on or collaborate with peers when they took this type of tests. That is, when taking tests for which marks were not assigned, they did not cheat. This suggested that cheating was a consequence of the pressure to obtain high marks.

The discussion in this sub-section indicates that the students' concerns about marks influenced their preferred assessment method (i.e., multiple-choice tests). They preferred multiple-choice tests because they thought that they could obtain higher marks by cheating and making random choices when answering test items. The concerns about marks also created test anxiety. They worried about obtaining low marks due to the difficulty of the tests. The students thought that their teachers and parents also were concerned about marks. This is the focus of the discussion in the next sub-section.

7.3.2. The students' perceptions of the teachers' and parents' concerns about marks.

Most students thought that their teachers and parents were concerned about marks. They believed that these stakeholders expected them to obtain high marks honestly.

7.3.2.1. The students' perceptions of the teachers' concerns about marks.

Many students thought that the teachers cared a great deal about their students' marks.

Interviewer: What do you think your teacher expects from you when she assesses you?
TB1S4: She expects many of us to obtain high marks.

When the students perceived that their teachers expected them to obtain high marks, they may try to obtain high marks by any means.

Some students thought that the teachers cared about their students' marks because they believed that their marks influenced how the teachers' teaching was judged. "She expects us to obtain good marks so that other people will make positive judgements about her teaching" (TA3S11). In the same vein, another student stated, "She expects us to obtain high marks so that other teachers compliment her" (TB3S3). These statements suggest that some students believed that their marks were important to the teachers because they influenced how other people judged their teachers' teaching.

A number of students thought that their teachers expected them to be honest in testing. "She expects us to take the tests using our knowledge and not to copy, no matter what marks we may obtain" (TB1S9). Other students shared a similar belief. "She wants us to do the tests by ourselves and not to copy. She wants real marks. She does not want high marks that are obtained by cheating" (TB3S4) and "She expects us to try our best and not to copy" (TB2S11).

Some students were well aware that their teacher tried to prevent them from cheating.

Interviewer:	In your opinion, why are the same items mixed to form 4 tests with 4 different codes?
TA1S3:	To prevent the students from copying.
TA1S1:	To prevent the students from copying.

The analysis indicated that some students perceived that their teachers were concerned about marks and they believed that the teachers expected them to obtain marks in a fair way.

7.3.2.2. The students' perceptions of parents' concerns about marks.

Most students believed that their parents expected them to obtain high marks from the tests because they perceived that their parents believed that marks were accurate measures of their children's competence.

Interviewer:	What about your parents? What do you think they expect from you in assessment?
Students:	Also high marks.
TA3S4:	Because they think marks reflect our learning outcomes =
TA3S3:	= our parents do not go to school, so they do not know how we take tests. They thinks marks are measures of our competence.

Similarly, another student stated, "High marks, of course, as high as possible, because they believe that my marks are representations of my effort and competence" (TB1S12). However, TA3S3 perceived that parents' belief that marks were accurate measures of their children's competence were naïve because "they do not know how we take tests."

In addition, some students believed that their parents expected them to obtain high marks and high rankings because these made their parents proud of them and provided status within their home communities.

> My parents want me to obtain high marks and high rankings. I think all parents want their children to obtain high marks and high rankings because their children's high rankings bring them the pride about their children. They feel proud of their children with their neighbours and friends. (TA2S6)

Another student's statement supported the assumption that students' high marks and rankings had some social value for their parents.

> I think not only my parents but other students' parents also expect their children to obtain high marks from tests and examinations. Last year, I was ranked as a good student. My

parents were very happy. They told their neighbours about my ranking. They felt very proud. (TB3S2)

Some students believed that their parents expected them to obtain high marks in an honest way.

Interviewer:	What do you think your parents expect from you when you take the tests?
TB3S5:	They expect me to do the tests by myself. . . . Marks have to reflect my real competence; they are not the results of cheating.
Interviewer:	What about your parents?
TB3S1:	They expect me to obtain high marks by myself.
Interviewer:	What about your parents?
TB3S4:	Just the same.
Interviewer:	And your parents?
TB3S2:	The same.

This sub-section indicates that the students perceived that both the teachers and their parents were concerned about marks. They believed that their marks and rankings influenced the teachers' and parents' professional and social status in their communities. However, the students thought that the teachers and their parents expected them to obtain high marks by fair means, not by cheating. The next section discusses the students' perceptions of the teachers' fairness and feedback.

7.4. The Students' Perceptions of the Teachers' Fairness and Feedback

The results of the analysis of the students' perceptions of assessment in their schools indicated that they distrusted the teachers' fairness. In addition, they thought that the teachers' feedback was not useful for improving their learning.

7.4.1. The students' distrust of the teachers' fairness.

Many students expressed distrust of the teachers' fairness in assessment. They thought that their teachers may construct tests that favoured those students who took private extra lessons with them.

Interviewer:	In your opinion, why does your teacher have to use the periodic tests that are designed by other teachers?
TA1S3:	So that none of the students will meet the items that they have met.
TA1S1:	The students often take private extra lessons with the teachers who teach them in school. If the teachers were allowed to design periodic tests for their own classes, those students who take private extra lessons with them would get more advantages.

Some other students held a similar belief.

Interviewer:	Do you know why the teachers are not allowed to design periodic tests for their own classes?
TB2S1:	The less competent students will attend the teachers' private extra lessons. If the teachers were allowed to design tests for their own classes, they may design tests in favour of these students.

Interviewer: What about you?

TB2S2: I share TB2S1's idea. This way of test design helps ensure fairness.

One student believed that the teachers constructed easier tests for their own students.

Interviewer: I am informed that the teachers in this school are not allowed to design the periodic tests for their own students. Do you know why?

TA3S7: Because there should be a common level for all the students in the same grade in the school. The teachers always want their own students to obtain higher marks than other teachers' students do.

The above statements suggested that some students believed that their teachers may advantage their own students.

A number of students thought that anonymous marking prevented the teachers from giving an advantage to some students when they marked the 45-minute tests and end-of-term tests.

Interviewer: Do you know why the periodic test papers of the students in different classes are mixed up and their names are cut off the test papers before the teachers mark them?

TA1S2: Because these tests have higher coefficients than the oral and 15-minute tests.

Interviewer: But why are they mixed up and why are your names not written on your test papers?

TA1S2: Because the teachers may favour their own students and the students who take private extra lessons in their homes.

Another student expressed a similar reason. "I think it is for fairness. Some teachers may be less strict when they mark their own students' test papers" (TB2S8).

This sub-section suggests that some students in both schools distrusted their teachers' fairness in constructing and marking the periodic tests.

7.4.2. The students' perceptions of the teachers' feedback.

Many students reported that the teachers' written feedback included only ticks, crosses, and marks. "The teacher rarely corrects our errors in the test papers. She points out correct and incorrect answers and she crosses the incorrect answers. That's all" (TA3S11). Some other students reported that their teacher's feedback was the same.

Interviewer: What do you see on your test papers when your teacher returns them to you?

TB1S11: Ticks and crosses in red ink.

TB1S10: That's all.

TB1S11: That's all. I see nothing else.

Some students thought that the teachers' feedback did not help them enhance their learning. One student expressed dissatisfaction with the teachers' oral feedback as follows.

He corrects our errors, but I think the correction is not thorough or detailed, so we do not know the nature of our errors. He just says that the answer is not that but this. Once I filled a gap with a word, and he crossed my word. I asked him, and he said that my answer was incorrect; the answer in the key was correct. He did not explain why my answer was

incorrect. Sometimes there are errors, and he does not explain our problems. He does not solve our problems (TA1S7).

Another student expected the teachers to show their students why errors were made. "Our teachers gave us just ticks and crosses and marks. We know which answers are correct and which are not, but we do not know why. I wish that they could explain our errors" (TB3S9).

This section indicates that some students valued feedback in helping them learn. However, they were unsatisfied with the teachers' feedback. They reported that teachers' written feedback included only ticks, crosses, and marks. Additionally, the teachers' oral feedback did not show them why errors were made and how these errors were avoided. They wanted feedback that helped them improve their learning.

7.5. Discussion

Four major findings can be drawn from the discussion in this chapter. First, the students perceived that assessment had multiple purposes. They perceived that their teachers used assessment as a means to monitor and control their learning. This perception was also held by the principals and EFL teachers in this study. This finding was similar to the findings in previous studies. Grade 8 students in Queensland (Moni et al., 2002) and New Zealand secondary students (G. T. L. Brown & Hirschfeld, 2008; G. T. L. Brown, Irving, et al., 2009) also thought that their teachers used assessment to monitor and control students' learning. Grade 8 students in Queensland perceived that assessment information was mainly used for report cards and that their teachers used assessment to monitor their learning (Moni et al., 2002). Congruent with policy-makers' regulations as documented in the policy documents examined in this study and the EFL teachers' perceptions, some students in the present study perceived that assessment was administered to obtain marks for ranking them. Indeed, obtaining marks for ranking students was the main purpose of the Vietnamese education system's policy-makers (MOET, 2011b, 2011f). This was also the main purpose that not only the students perceived, but that the EFL teachers both perceived and enacted.

Additionally, the students in the present study believed that assessment informed themselves and their teachers about their learning. They also thought that assessment informed teachers about their teaching. The students thought assessment information could be used to modify and improve teaching and learning. These formative purposes of assessment were expected by policy-makers and also perceived by the principals and EFL teachers in the present study. School students in previous studies also perceived these improvement purposes of assessment. For example, Grade 8 students in Queensland thought that assessment informed teachers about their learning and helped them improve it (Moni et al., 2002). New Zealand secondary students perceived that assessment helped

them improve their learning through teachers' feedback (Peterson & Irving, 2008) and helped teachers and students improve teaching and learning (G. T. L. Brown, Irving, et al., 2009).

The second major finding was related to the students' concerns about marks and their perceptions of the teachers' and parents' concerns about marks. Specifically, most students thought that marks were not accurate measures of their learning outcomes. This perception was consistent with a perception held by Grade 8 students in Queensland, Australia (Moni et al., 2002) and high school students in Turkey (Yildirim, 2004). This perception was also consistent with the principals' and EFL teachers' reports that teachers constructed tests, especially continuous tests, according to teachers' expectations of their students and students' competence rather than according to standards of knowledge and skills as expected by policy-makers and the principals. In addition, the marks were not accurate measures of students' learning outcomes because the EFL teachers reported that they modified the difficulty of their tests so that their students could obtain high marks and that they used marks to reward and punish their students in order to control their learning. In addition, some students reported that they cheated when taking tests in order to obtain good marks. In contexts where marks are high stakes for students and teachers, both students and teachers may try to raise students' marks in inappropriate ways.

The majority of students in the present study reported that they were concerned about marks. This finding was consistent with the EFL teachers' perceptions that the greatest concern of their students in assessment was marks. Students are concerned about marks even when they are not high-stakes to them. In Smith and Gorard's (2005) study, most Grade 7 students in a secondary school in Wales wanted to be given marks because they believed that marks informed them about their performance (Smith & Gorard, 2005). Some Grade 9 and 10 students in a study in New Zealand also cared about marks rather than the comments given by their teachers because "at the end of the day the grade gets you passed anyway" (Peterson & Irving, 2008, p. 246). Because they were concerned about marks, the students in the present study preferred multiple-choice tests, though they did not believe that these tests accurately measured their learning outcomes. They preferred multiple-choice tests because it was easier for them to cheat, to guess correct answers, or simply to tick responses to the items randomly. They reported that they preferred multiple-choice tests to open-ended and oral tests because they did not have to face their limited command of English when they wrote and spoke English. High school students in Turkey also preferred multiple-choice tests in social studies because they did not have to construct responses by themselves and they could guess correct answers (Yildirim, 2004). In addition, they wanted to prepare for the university entrance examination, which included multiple-choice questions (Yildirim, 2004). The finding in this study and that of Yildirim (2004) suggested that when

assessment results were high-stakes, students were more concerned about obtaining good marks than demonstrating their knowledge and skills and learning from assessment.

Most students in the present study reported that they were concerned about marks because they perceived that marks had important consequences, and that marks were used to make judgements about their learning and to rank them. They believed that marks influenced their image in other people's mind and their parents' reactions to them (i.e., giving them rewards or scolding them). They also perceived that marks were part of the competition among students in learning. Indeed, as perceived by the EFL teachers, marks not only had consequences for students but also for teachers and parents. Grade 7 students in Wales perceived that marks motivated them to obtain higher marks and marks were necessary for them to inform their parents about their learning (Smith & Gorard, 2005). Some Grade 9 and 10 students in New Zealand associated marks with their pride, and they reported that they compared their marks with peers (Peterson & Irving, 2008). The findings related to the concerns about marks of the students in the present study and in previous studies suggested that students in different contexts were concerned about marks because they perceived that marks had important functions and consequences.

A great number of students in the present study reported that they felt anxiety because of the pressure to do well in testing. High school students in Turkey, for example, also reported that they felt anxious in testing (Yildirim, 2004). Grade 3 to 6 students in the US expressed anxiety when they were asked to draw pictures and written descriptions of their drawings after they had finished their high stakes tests (Triplett & Barksdale, 2005). Similarly, Grade 9 and 10 students in New Zealand reported that their parents' expectations created pressure to obtain good marks (Peterson & Irving, 2008). These findings suggested that test anxiety was common in different contexts and it was partly created by the pressure to do well in testing.

Some students believed that their teachers expected them to obtain high marks because their marks influenced how the teachers' teaching was judged. This finding was consistent with several teachers' report that their students' marks influenced how their teaching was judged. Similar to the students in this study, some students in a study in the US perceived that their teachers expected them to obtain high marks (Brookhart & Bronowicz, 2003). These findings indicated that students in different contexts thought that their teachers expected them to obtain high marks. The finding in the present study suggested that when students perceived that their marks had consequences for their teachers, they thought that their teachers expected them to obtain high marks.

Most students in the present study perceived that their parents were concerned about marks. Many students thought that their parents expected them to obtain high marks because their parents believed that marks reflected their learning and influenced the parents' pride in their children with

neighbours and friends. Parents' concerns' about their children's marks were also reported by the EFL teachers in this study. Some teachers thought that parents were concerned about their children's marks because parents believed they influenced their social status. Several students said that their parents' expectations of high marks was part of the pressure that made them cheat when taking tests and examinations.

Third, some students expressed distrust of the teachers' fairness in assessment. They believed that when constructing classroom tests and marked test papers, some teachers may favour those students who paid to attend extra private lessons in the teachers' homes. The cause for the students' distrust of their teachers' attempts at fairness was similar to the cause of the principals' distrust of their teachers' abilities to be fair. Some high school students in Queensland also expressed distrust of teachers' fairness, and they thought that their teachers assigned marks according to their personal feelings about students (Moni et al., 2002).

Fourth, some students were dissatisfied with the teachers' feedback on their errors. They expected feedback that showed them how to improve their learning, but they reported that the teachers' written feedback included only ticks, crosses, and marks. They thought that both written and oral feedback were not helpful for improving their learning. In the classroom lessons, the EFL teachers provided their students with direct corrective feedback and feedback about the students as persons. Similar to the students in this study, Grade 9 and 10 students in New Zealand reported receiving ticks, crosses, marks and sometimes comments including "constructive criticism", "tips", or "reminders" (Peterson & Irving, 2008, p. 246). The students in New Zealand also expected feedback that showed them why they were wrong and how they could improve their learning. Likewise, secondary students in Hong Kong perceived that their teachers' feedback was not specific enough for them to see their strengths and weaknesses (Lee, 2008). The Hong Kong students expected to receive not only written feedback but also oral feedback, class discussions, and conferences with teachers. The finding about the perception of teachers' feedback held by the students in the present study and those in previous studies indicated that students in different contexts were not satisfied with their teachers' feedback, and they expected feedback that could help them improve their learning. The reason for the EFL teachers in the present study to give their students only two types of feedback may be because the teachers were not given guidelines about how to use assessment for formative purposes.

This chapter discussed the students' perceptions of assessment in their contexts. The next chapter reports the assessment practices that were observed in the EFL teachers' classrooms.

Chapter 8. Assessment Practices as Observed in the EFL Classrooms

8.1. Introduction

This chapter presents the results of the analysis of the classroom observations undertaken in order to explore the EFL teachers' assessment practices in their classrooms. The results of the analysis are set out in Figure 5. This chapter is organised according to the themes and theme clusters in Figure 5. The first cluster is about the assessment practices that the EFL teachers carried out in their classrooms. The first theme in this cluster indicates that the teachers relied on the tasks and exercises in the textbooks that they used in their classrooms. These tasks and exercises mainly assessed the students' recall of factual information, vocabulary, and grammar rules. In addition, the teachers used quizzes in their classrooms and they used them to assess the students' recall of vocabulary. The classroom observations indicated that the teachers used oral questions and observations as formative assessment. However, the analysis of these strategies was not presented in separate themes or categories because these strategies were viewed as aspects of the teachers' assessment of the language skills, grammar, vocabulary, and pronunciation. The second theme reveals that the teachers mainly provided direct corrective feedback to the students. The teachers also used praise for the students as persons.

The second cluster is related to ethical issues in assessment in the observed classrooms. The first theme is about how the teachers ensured fairness in assessment. The second theme presents how the teachers used marks to reward and punish their students.

Figure 5. Results of the Analysis of the Classroom Observations

Assessment Practices as Observed in the EFL Classrooms		
Assessment Practices in the Classrooms	**Ethical Issues in Assessment**	Clusters
Using the tasks and exercises in the textbooks and teacher-developed quizzes to assess lower-order thinking skills - Reliance on the tasks and exercises in the textbooks - Teacher-developed quizzes for assessing recall of vocabulary **Feedback on the students' errors** - Direct corrective feedback for verification of the students' errors - Praise as feedback about the students as persons	**Ensuring fairness in classroom testing** - Ensuring fairness by providing equitable conditions for testing - Preventing cheating in testing **Marks as rewards and punishment** - Using marks as extrinsic rewards for voluntary participation - Using marks to punish the students for not doing homework	Themes and Categories

Figure 5. Categories, themes, and theme clusters emerging from the analysis of the 42 classroom observations

8.2. Assessment Practices in the Classrooms

The results of the analysis of the classroom observations indicated that the teachers relied on the tasks and exercises prescribed in the lessons on language skills (i.e., listening, speaking, reading, and writing), grammar, vocabulary, and pronunciation in the textbooks. Assessment in the observed lessons focused on lower-order thinking skills. The feedback given in the observed classrooms included direct corrective feedback and praise about the students as persons.

8.2.1. Using the tasks and exercises in the textbooks and teacher-developed quizzes to assess lower-order thinking skills.

The teachers relied on the tasks and exercises in the textbooks to assess their students. In addition, teacher-developed quizzes were used in some of the observed lessons. However, the teachers used the tasks, exercises, and quizzes mainly to assess lower-order thinking skills. Specifically, they assessed the students' recall of facts, vocabulary, and grammar rules. They also assessed the students' comprehension.

8.2.1.1. Reliance on the tasks and exercises in the textbooks.

The EFL teachers in high schools in Vietnam used the same textbooks for mainstream classrooms. These textbooks were written by the MOET. The English textbooks for the 10th, 11th, and 12th Grades had the same structure. Each textbook consisted of 16 units. Each unit covered one topic (e.g., "Special education" and "The mass media" in *English 10*). Each unit consisted of five lessons. The first lesson in each unit focused on reading; the second, speaking; the third, listening; the fourth, writing; and the fifth, called Language Focus, grammar, vocabulary, and pronunciation. Each lesson on listening, speaking, reading, and writing included two or three "tasks". Each Language Focus lesson consisted of three or four "exercises". Because researchers as well as teachers have different understandings of what a "task" in language teaching is (Ellis, 2009a; Nunan, 2006), in this chapter the words "task" and "exercise" are used as they appeared in the textbooks.

The teachers used the tasks and exercises in the textbooks as assessment activities in their classrooms, even though the principals, the teachers, and the students did not mention them as assessments during interviews. The structure of each unit in the textbooks also required that all language skills, grammar, and vocabulary be assessed in the classrooms. Following are some examples of how the teachers used these tasks and exercises as assessments of these language skills.

In the following example, a teacher in School B was working with her students to complete Task 1 (p. 150) of *English 11*. The students were asked to listen to the tape and decide which statements about the hobbies of the speaker in the tape were true and which ones were false.

> The teacher played the entire talk, then she asked her students, "How many questions can you answer?" Many students said that the talk was too difficult for them to understand. The teacher said, "OK, now I play just a short part of the tape. Listen and see whether the first statement is true or false. What do you listen for in order to answer question 1?" Some student said, "His hobby." The teacher asked, "His hobby or the time when he started his hobby?" Some student said, "The time, not his hobby." The teacher said, "That's right. Pay attention to the words about time. Now, listen."
>
> The teacher played a short part of the tape and asked the students, "When did the student start his hobby?" Some students said, "When he was young." The teacher said, "So, the first statement is . . . ." Some students said that the first statement was true. The teacher said, "Very good. The first statement is true." . . .
>
> The teacher used these procedures to elicit the students' answers to the other seven statements in Task 1.

<div align="right">(TB1Ob.L)</div>

In this example, the teacher assessed the students' understanding of what they listened to and their ability to remember the details using true/false questions. These assessed the students' lower-order thinking skills, i.e. comprehension and recall of what they heard from the tape. The teacher also tried to support the students by giving them clues about what they listened for using key words in the questions.

The teachers also used the tasks in the speaking lessons in the textbooks as assessments. For example, in a speaking lesson in School A, a teacher told the students to read a short passage in Task 1 (p. 126) of *English 10* and answer the questions below the passage. The instructions and the questions were as follows.

> *Work in pairs*. Answer the questions.
> 1. What kind of music does Ha Anh like?
> 2. Why does she listen to it?
> 3. What is her favourite band?
> 4. When does she listen to music?

<div align="right">(*English 10*, p. 126, emphasis in original)</div>

In this passage, a girl named Ha Anh wrote about her favourite music ("Yes, I love pop music"), why she loved pop music ("It keeps me happy"), her favourite music band ("I really like the Backstreet Boys"), and when she listened to music ("I listen to music when I'm out"). This task was designed for students to work in pairs to practice speaking English using a short text and the guiding questions. However, the teacher used the task as an oral test to assess a student's lower-order reading skill in literal comprehension. Specifically, the teacher checked the student's ability to identify facts in the text when she asked a student to answer the questions as described in the following extract.

After three minutes, she asked a student to stand up and asked him these questions. The student answered all her questions correctly. The teacher complimented the student and asked him to sit down.

(TA2Ob.S)

Instead of asking the students to work in pairs as suggested in the textbook, the teacher asked only one student to answer the questions orally, thus turning an activity intended to help students practice speaking into an assessment of one individual student's recall of factual information and assessed reading skills.

The classroom observations indicated that assessment of vocabulary focused on the decontextualized meanings of words and their pronunciation. Assessment of English vocabulary in this way was seen in nearly every lesson of the EFL teachers in both schools. In the following example, a teacher in School B assessed her students' memory of the Vietnamese equivalents and the pronunciation of some English words that she had taught in the previous lesson.

The teacher entered the classroom. She greeted the students in response to their greeting. She sat down at her desk. She opened her mark book. The teacher said: "Last time I told you to learn by heart the words that we learned in the reading lesson. Did you learn the words? Does anyone want to write these words on the blackboard and pronounce them?". None of the students in the class raised his or her hand.

The teacher pointed to a student sitting in the middle of the class and then another student at the back of the class, saying, "You. Go to the blackboard. And you."

The students went to the blackboard. The teacher said, "Stand on the left. And you, on the right." She explained that she would read the Vietnamese words and the students would write the English equivalents on the blackboard. She asked Lan (pseudonym) to write the English equivalents of the following Vietnamese words on the left-hand side of the blackboard. These words were in the reading lesson on page 105 of *English 10*.

Tàn phá [destroy]
Sự đa dạng [variety]
Loại bỏ [eliminate]
Liên tục [constant]
Sự tuần hoàn [circulation]

Then she asked Nam (pseudonym) to write the English equivalents of the following Vietnamese words on the right-hand side of the blackboard.

Có giá trị [valuable]
Mùa [season]
Môi trường [environment]
Đe dọa [threaten]
Thông qua [pass]

Lan was able to write the English equivalents for the first four words, but she could not write the last word. Similarly, Nam was able to write the first four words but not the last word.

The teacher asked Lan and then Nam to read aloud the words they had written on the blackboard. Lan pronounced the word "variety" incorrectly, and Nam pronounced the word "threaten" incorrectly.

The teacher then asked two other students to pronounce the word "variety", but they did not pronounce it correctly either. A fifth student stood up and voluntarily pronounced it correctly and the teacher said, "Very good. That's correct pronunciation." Then she asked

another student to pronounce the word "threaten", but he did not pronounce it correctly. A girl sitting at the front of the class volunteered to pronounce the word, and she pronounced it correctly. The teacher said, "Well done. I give Lan and Nam 7 marks each. Vocabulary is very important. You have to remember Vietnamese equivalents and you have to pronounce English words correctly. Every day, learn some English words."

(TB2Ob.S)

This extract indicates that the teacher valued her students' memory for writing English words and being able to pronounce . Also, she relied on the textbook, specifically the reading lesson, when she assessed her students' vocabulary and pronunciation.

The teachers also used the tasks in the reading lessons as assessments. In the following example, a teacher in School B used Task 2 (p. 164) of *English 12* as an assessment. This task asked the students to read a passage about "Women in society" (p. 163) and answer five multiple-choice questions (p. 164). The students worked individually while the teacher walked around the classroom monitoring them.

After seven minutes, the teacher asked the students, "Which answer do you choose for question 1? According to the passage, what was the main role traditionally accorded to women?" The students called out 3 different answers (i.e., A, B, and C). The teacher said, "The correct answer is C. What words help you find the answer to this question?" Some student said, "Main." Another student said, "I think 'traditional.'" The teacher pointed at a student and asked, "Which word?" The student replied, "I think 'role.'" The teacher said, "Yes, 'role.'" She read the first paragraph of the passage on page 163 and explained why C was the correct answer. . . .

Similar procedures were used to provide the students with the answers to the other four questions in Task 2.

(TB3Ob.R)

This teacher tried to elicit the answers to the questions in the textbook. She gave the students direct corrective feedback (i.e., providing the students with correct answers). In addition, she taught the students how to use key words in the questions as clues for skimming for details in the passage. Therefore, she integrated assessment into her teaching. However, her assessment focused on the students' low-order thinking, and she provided her students with only direct corrective feedback.

The teachers used the tasks in the writing lessons to assess writing. In the following episode, a teacher in School A used Task 2 (p. 158) of *English 12* as a writing test. This task required the students to write a paragraph of about 100 words explaining the reasons why they would like to work for WWF, WHO, or UN. The students were asked to start their paragraph with "*I'd like to work for the …….. for a number of reasons. First, ……….*".

The teacher asked the students to write the paragraph on a piece of paper. The students worked individually while the teacher was sitting at his desk. After 10 minutes, he told the students to write their name on their paper and be ready to hand in the paper for him to mark later in place of a 15-minute test.

(TA1Ob.W)

In this case, the teacher did not tell his students that he would collect the students' papers and assign marks. Perhaps he had intended to collect the papers and assigned marks, so asked the students to work individually and he did not provide the students with any support.

The teachers also assessed the students' ability to recall grammar rules. In the following extract, a teacher in School A checked the students' abilities to remember the passive voice structure in English, which they had learned in the previous lesson.

> After greeting the students, the teacher took a seat at his desk. He asked a student to write the formula of the passive voice structure on the blackboard. The student wrote the following formula on the blackboard:
> S + V + O => S + TO BE + PP + by + O
> The teacher asked the whole class, "Is the formula right or wrong?" Many students said, "Right. Right." The teacher wrote the following sentences on the blackboard:
> He bought a new motorbike last week.
> We will finish our lesson in about 40 minutes.
> Then he asked the student to change the sentences into the passive voice. The student wrote the following sentences on the blackboard:
> A new motorbike was bought by him last week.
> Our lesson will be finished in about 40 minutes.
> The teacher asked the students, "Are they correct?" One student stood up and said, "I think we have to say 'the previous week or the week before, not last week.'" Some other students said, "Yes. The previous week or the week before." The teacher said, "Yes, you are right." He said to the student who was asked to write the formula and to change the sentences into passive sentences, "Good. Eight marks."

<div align="right">(TA1Ob.R)</div>

This example indicated that the teacher wanted to see whether or not the student remembered the passive voice structure in English and whether or not the student could apply this grammar structure. In addition, the teacher tried to involve the students in the assessment activity by asking them to give judgements about their peers' answers. Moreover, the teacher expected his students to reproduce textbook answers. In fact "last week" could be correct. The context of the sentence "He bought a new motorbike last week" was not specific enough for the students to decide which answer, "last week" or "the week before" or "the previous week", was correct. In this case, the teacher was not flexible to accept a possible alternative and take advantage of the emerging situation to improve his students' knowledge.

The classroom observations reported in this sub-section indicated that the teachers relied on the tasks and exercises in the textbooks. They used these tasks and exercises to assess the students' recall of factual information, vocabulary, and grammar rules. The teachers were not flexible in using assessments from other sources for classroom assessment. They adhered themselves to the tasks and exercises in the textbooks and textbook answers. The teachers involved the students in classroom assessment. However, the students' feedback as well as the teachers' feedback to the students' errors was mainly direct corrective feedback.

8.2.1.2. Quizzes for assessing recall of vocabulary.

Thinking skills have been arranged in hierarchical order in some taxonomies, with "knowledge" in the lowest position, and "evaluation" in the highest position, and "comprehension", "application", "analysis", and "synthesis" in between (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956). Although all these thinking skills need to be developed in education, the higher-order and middle-order thinking skills (i.e., application, analysis, synthesis, and evaluation) are more desirable than the lower-order thinking skills (i.e., knowledge and comprehension).

However, some teachers in both schools used quizzes to assess the students' abilities to recall vocabulary. A quiz typically began when the teachers organised the class into two teams and told the students in the teams to provide an English word when they were given a picture as a clue. In some quizzes, the students were asked to provide an English word when they saw a Vietnamese word flash on the TV screen or the other way round.

In the following example, a teacher in School B assessed the students' recall of the English names of the countries in ASEAN (The Association of Southeast Asian Nations). These names had been taught in the previous lesson. After the teacher organised the class into two teams, she said that she would show the national flags of 10 ASEAN countries on the TV screen. The first student to raise his or her hand and provided the correct name of a country when a flag flashed on the TV screen would score one point for his or her team.

> After telling the students the rules of the quiz, the teacher showed the national flags of Vietnam, Malaysia, Thailand, Indonesia, the Philippines, Singapore, Brunei, Myanmar, Laos, and Cambodia. Each flag appeared on the TV screen for about one second. The students were very excited, and the classroom was very noisy. They tried to be the first to call out the names of the countries. Finally, both teams scored five points each.

(TB3Ob.S)

This example indicated that the students were actively involved in the quiz. They were taking part in the classroom assessment and enjoyed some fun at the same time.

This sub-section indicates that the teachers mainly used the tasks and exercises in the textbooks and teacher-developed quizzes as assessment activities. However, these classroom assessments were mainly used to assess the students' lower-order thinking skills, specifically their comprehension and recall of factual information, vocabulary, and grammar rules. The teachers' focus on lower-order thinking skills was influenced by the tasks and exercises in the textbooks. It may also be affected by policy-makers' expectations, which required that tests included knowledge questions, comprehension questions, and application questions (see Chapter 5). This influence was possible because "what amount and what kind of higher-order thinking should be required on a classroom assessment depend on the particular learning goals to be assessed" (Brookhart, 2010, p.

2). The teachers' focus on recall was a common practice because recall-level questions are easy to ask and they are the most common questions asked by even the best teachers (Brookhart, 2010).

8.2.2. Feedback on the students' errors.

Feedback is an integral part of assessment practices, and the extracts in the previous sections indicated that the teachers provided their students with feedback in their classroom assessment. However, the results of the analysis of the classroom observations indicated that the teachers' feedback was not helpful for learning improvement as expected by the students. The classroom observations indicated that the feedback given by either the teachers or peers in the observed lessons mainly included direct corrective feedback on the students' errors and feedback about the students as persons.

8.2.2.1. Direct corrective feedback for verification of the students' errors.

Feedback, defined as "any information that is provided to the performer of any action about that performance" (Black & Wiliam, 1998a, p. 53), "plays a central role in supporting and promoting students' language learning" (Inbar-Lourie, 2008a, p. 387) and is "one of the most powerful influences on learning and achievement" (Hattie & Timperley, 2007, p. 81). It is viewed as an essential part of formative assessment and a means to promote and enhance learning (Butler & Lee, 2010; Dann, 2002; Davison & Leung, 2009; Elwood & Klenowski, 2002; Inbar-Lourie, 2008a).

Researchers have indicated specific qualities of useful feedback. For feedback to be useful for promoting and enhancing learning, it has to be timely, specific, understandable, and actionable (Fisher & Frey, 2011). Quality feedback has to show students what to do to improve their learning (Elwood & Klenowski, 2002; Shute, 2008b). Nicol and Macfarlane-Dick (2006) suggest that good feedback 1) shows students what good performance is, 2) helps develop self-assessment in learning, 3) informs students about their learning, 4) encourages teacher-student and student-student dialogues about learning, 5) encourages positive beliefs and self-esteem, 6) helps students close the gap between current and desirable performance, and 7) provides teachers with information that they can use to modify teaching (Nicol & Macfarlane-Dick, 2006).

There are various types of feedback. Feedback can be as simple as scores, letter grades, or ticks and crosses or as complicated as "models, cues, prompts, hints, partial solutions, and direct instruction" (Shute, 2008a, p. 163). Feedback can be written or oral, and given either by teachers or peers (Frey & Fisher, 2011). Feedback can be related to four different "levels" (Frey & Fisher, 2011, p. 64): feedback about the task, the processing of the task, self-regulation, and the self as a person. Feedback about the task informs the student about the quality of his or her performance on

the task. Feedback about the processing of the task shows the student the processes she or he needs to use to complete the task. Feedback about self-regulation is related to the student's self-appraisal and self-management. That is, it shows the student how to assess his or her ability, knowledge, behaviours, and actions in order to achieve his or her goals. Feedback about the self as a person focuses on the student himself or herself. It is related to the student's effort, interest, engagement, and efficacy. Praise belongs to this type of feedback. Feedback can be categorised into facilitative feedback (i.e., providing cues and guidance) and directive feedback (i.e., providing corrections) (Shute, 2008b). Feedback can be evaluative (i.e., providing judgements about the student's performance on the task) or descriptive (i.e., providing cues and guidance on how to correct errors) (Davis & Neitzel, 2011).

The most common type of feedback is called corrective feedback (Airasian, 1997). This type of feedback focuses on the task (Frey & Fisher, 2011). Specifically, corrective feedback corrects misunderstandings or incorrect answers (Frey & Fisher, 2011). With corrective feedback, the teacher identifies correct or incorrect answers, asks students to provide either additional or different information, and draws students' attention to specific knowledge (Frey & Fisher, 2011). According to Ellis (2009b), corrective feedback includes direct corrective feedback and indirect corrective feedback. With direct corrective feedback, the teacher identifies an error and provides the correct answer. With indirect corrective feedback, the teacher just indicates an error without providing the student with the correct answer.

Different types of feedback have different values for student learning. For example, corrective feedback about the task is useful when it is used to address mistakes, while feedback about self-regulation helps students regulate their actions and behaviours (Fisher & Frey, 2011). In addition, students at different stages of learning respond and benefit differently from different types of feedback. For instance, directive feedback (i.e., providing corrections) is believed to be more helpful in the early stages of learning, while facilitative feedback is said to be more helpful in the later stages of learning (Shute, 2008b). This suggests that feedback of various types should be provided for different students to improve their learning.

However, the classroom observations in this present study indicated that feedback on the students' errors included only direct corrective feedback. Following are two examples illustrating where the teachers provided only this type of feedback to the students' errors. The first example was an assessment in a "language focus" lesson in which a teacher in School B worked with the students to complete Exercise 1 (p. 139) of *English 10*.

> The teacher asked two students to go to the blackboard to write the present participles and past participles of the verbs given on page 139 of *English 10*.
> 1. fascinate   …………………….. 6. bore   …………………….

2. excite	……………………..	7. surprise	………………….
3. terrify	……………………..	8. amuse	………………….
4. irritate	……………………..	9. embarrass	………………….
5. horrify	……………………..	10. frustrate	………………….

Two students volunteered to go to the blackboard. The first student, who wrote the present and past participles of the first five verbs, made two errors (i.e., terrifyed and horrifyed). The teacher asked the other students in the classroom to comment on the participles written on the blackboard.

One student said that the past participles of "terrify" and "horrify" were wrong and that the correct forms were "terrified" and "horrified".

The teacher changed "terrifyed" and "horrifyed" into "terrified" and "horrified". Then he told the students to write the participles in their notebooks.

(TB2Ob.LF)

In this case, the teacher asked the students other than the one who made the errors to indicate the incorrect answers and provide correct ones. The teacher did not explain why "terrifyed" and "horrifyed" were incorrect. Nor did she ask the students who provided the correct answers to justify their answers. This practice was typical of assessment of learning.

In a typical test-return lesson observed in the classrooms, the teacher and his or her students provided correct responses to the items in a multiple-choice test. In the following example, a teacher in School A returned test papers back to the students and provided the correct responses or asked the students to provide the correct answers to 20 multiple-choice items in the test.

The teacher gave some comments on the students' marks obtained from a 15-minute test that they took in the previous week. She said that many students got high marks and that she was pleased with their marks. Then she returned the test papers back to the students. She told the students to have a look at their test paper and then work with her to give correct responses to the items in the test.

After 5 minutes, she asked one of the students, "Now, for question 1, which answer is correct?" The student said that C was the correct response. The teacher asked, "Everyone, is C the correct answer?" Many students said that C was the correct response. The teacher agreed that C was the correct response and she wrote C on the blackboard.

The teacher asked, "What about question 2? Which answer is correct?" and she pointed at a student sitting at the back of the classroom. The student stood up and said, "The correct answer is A." The teacher said, "Right. Very good" and she wrote A on the blackboard.

She asked, "Now, question 3. Which answer?" She pointed at a student sitting in the front. The student replied, "B. I think B." The teacher shook her head. Some students said, "C. C is the correct answer." The teacher pointed at another student. The student stood up and said, "The correct answer is C." The teacher said, "Good" and she wrote C on the blackboard. . . .

Similar procedures were used to provide the correct answers to the other questions in the 15-minute test.

(TA3Ob.TR)

In this excerpt, the teacher either verified whether or not a response was correct and then provided a correct answer or she asked the students to do so. The teacher did not provide feedback that the students could use to enhance their performance when they met questions similar to the ones in the

test (Davis & Neitzel, 2011). Nor did the teacher provide feedback that helped the students improve their work (Harlen & Winter, 2004). In addition to direct corrective feedback, the teachers often praised their students.

8.2.2.2. Praise as feedback about the students as persons.

According to Frey and Fisher (2011), teachers' feedback may focus on four objects: (a) the errors that the student has made when completing a task, (b) the processes that the student uses to complete a task, (c) what the student needs to regulate to achieve a learning goal, and (d) the student as a person. Frey and Fisher say that praise belongs to the last type of feedback. They argue that praise is prevalent in the classroom but not effective for learning improvement.

In these classrooms, the teachers often used praise as feedback to the students. Following is an illustration of the prevalence of this type of feedback.

> The teacher told her students to look at the TV screen and give the Vietnamese equivalents of the English words when they were invited to do so.
> She showed the word "guitarist" on the TV screen and pointed at a student at the back of the classroom. The student stood up and called out the Vietnamese word (người chơi đàng guitar). The teacher said, "Correct. Very good. Sit down please."
> Next, she showed the word "tune (in music)" and pointed at a student sitting on the left of the classroom. The student stood up and said, "Giai điệu." The teacher said, "Very good. Sit down, please."
> The teacher showed the word "stamp collector" on the TV screen and asked, "Who know this word?" Many students raised their hands. The teacher pointed at a student sitting at the front of the class. The student said, "Người sưu tập tem." The teacher asked the whole class, "Correct, everyone?" Many students shouted, "Yes. Correct, teacher." The teacher said, "Good. Sit down, please."
>
> (TB1Ob.R)

This example showed that the teacher verified the students' answers using direct corrective feedback and praise at the same time (e.g., "Correct. Very good. Sit down please."). The teacher also used praise alone to indicate that the student's answer was correct (e.g., "Good. Sit down, please."). Praise such as "Good" and "Very good" may have a positive impact on students' attitudes towards learning, but it has a limited effect on students' learning per se (Black & Wiliam, 1998a). However, this type of feedback was frequently used in the classrooms of the participant teachers.

The analysis up to this point indicates that the teachers relied on the tasks and exercises in the textbooks for classroom assessment. They mainly assessed the students' lower-order thinking skills. Specifically, they focused on the students' abilities to recall facts, vocabulary, and grammar rules. Feedback was found to be provided in all the observed assessments. However, feedback included only two types: direct corrective feedback and praise. They provided direct corrective feedback when they verified an error and provided a correct answer. They used either praise with direct corrective feedback or praise alone as a signal verifying the students' answers were correct.

In a meta-analysis of the literature on feedback, Hattie and Timperley (2007) found that different types of feedback have different effects on student learning. They also found that the most effective feedback is the feedback that informs students about their performance on a task and how to do it more effectively. Praise, rewards, and punishment are the least effective feedback. Therefore, the feedback provided by the teachers in the present study may not be useful for student learning. In effect, students in this present study thought that their teachers provided them only marks, ticks, and crosses, and they thought that the teachers' feedback did not help them improve their learning.

The analysis of the classroom observations in this section provides some understandings about the types of assessment implemented in the teachers' classrooms and the types of feedback provided to the students. A number of ethical issues also emerged from the analysis of the classroom observations. These issues are discussed in the following section.

8.3. Ethical Issues in Assessment

The classroom observations indicated that the teachers tried to ensure fairness in testing by providing equitable conditions when they implemented 15-minute tests in their classrooms. In addition, they prevented cheating that occurred in some classrooms when the teachers implemented these tests. The classroom observations also indicated that some teachers used marks as rewards, threats, and punishment in their classrooms.

8.3.1. Ensuring fairness in classroom testing.

When administering the 15-minute tests, the teachers tried to ensure that all the students took the test in the same conditions. Moreover, they tried to prevent their students from cheating when they implemented these tests.

8.3.1.1. Ensuring fairness by providing equitable conditions for testing.

For each teacher, the researcher observed one lesson in which the teacher implemented a 15-minute test. These were mandatory 15-minute tests that the teachers implemented to generate marks. In these tests, ensuring equitable conditions for all the students in the classroom was an integral part of the teachers' work. The following extract showed typical procedures that the teachers used in a 15-minute test.

> The teacher said to the whole class, "Last time I told you that we will take a 15-minute test today. Did you revise your lesson for this test?" The students made a lot of noise. Some said yes and many said no.
> Then the teacher told the students to put all of their books, notebooks and any paper into the compartment in their desks. She asked the students to keep some distance from each other. Four students sat at each desk with a length of about 2 metres. When no books and

notebooks were seen on the students' desks and the students sat in straight lines, the teacher asked them not to read the test until she told them to start.

> The teacher asked the monitor to hand out the test papers to the students while she stood at the front of the classroom looking at the students and reminding them not to read the test papers. When the monitor had handed out all the test papers, the teacher told the students to start the test and she wrote "9:50" in the upper right corner of the blackboard.
>
> . . .
>
> At 10am, the teacher said, "You have only 5 minutes left." At 10.05am, the teacher told the students, "Stop writing. Put down your pen on your desk. Put your paper at the end of your desk. I will subtract 2 marks if you continue writing."
>
> The teacher asked the monitor to collect the students' test papers while she stood at the front of the classroom looking at the students. The monitor gave her the test papers, and the teacher counted these test papers. Then she said, "There are 44 test papers. There are no absentees today. You were all very good today. Nobody was caught cheating." The students laughed.

(TA2OB.T)

Fairness in testing is an aspect of ethics (Elwood, 2013), and providing equitable conditions to all test-takers is necessary to ensure fairness in testing (Tierney, 2014). The above excerpt indicated that TA2 tried to ensure that all the students in the classroom took the test in the same conditions. This way of ensuring fairness in testing was necessary because test results were used to rank the students and to promote or retain them.

The excerpt also showed that TA2 also tried to prevent the students from cheating by asking the students to put all their books, notebooks, and any paper into the compartment in their desks. In addition, TA2 asked the students to keep some distance from each other by asking them to sit in straight lines. Her statement at the end of the test (i.e., "You were all very good today. Nobody was caught cheating") may be just a joke, but it may also be interpreted that no cheating in that test was an exception rather than the norm.

8.3.1.2. Preventing cheating in testing.

Prevention of the students' cheating was seen in 15-minute tests in both schools. The students did not refer to the textbooks or any written materials because the test procedures prevented them from cheating in this manner. However, cheating was observed during several 15-minute tests. Cheating occurred in the form of oral exchanges between the students sitting near each other. The following excerpt illustrated how a teacher in School B stopped her students from cheating during a 15-minute test.

> Five minutes after the teacher told the students to start the test, two students sitting on the left side of the classroom were whispering to each other. The teacher knocked her big ruler on her desk and said, "Stop! No talking. You are doing a test." The classroom became quiet again.
>
> Four minutes later, two students sitting at the back of the classroom whispered to each other. The teacher left her chair and walked to the whispering students. She said, "Stop

cheating or else I will subtract two marks each when I marks your papers." The students scratched their heads.

The teacher announced that only 5 minutes was left for the students to answer the test questions. The classroom became noisy. A lot of whispers could be heard across the classroom. The teacher knocked her ruler very hard on her desk. She said, "Stop! Stop! No more talk." The classroom became quiet again.

Then, the teacher told the students to stop writing because the time for the test was over. Some students bargained with the teacher. They asked the teacher to give them some more time because the test was difficult. The teacher said, "No more time. Stop writing and put your test papers at the end of your desk."

She asked the monitor to collect the test papers while she stood at the front of the classroom monitoring the students to make sure that none of them continued writing.

The monitor gave the teacher the test papers. The teacher counted the test papers, while the students asked each other the answers to the questions on the test.

(TB2Ob.T)

The above excerpt indicated that more students whispered to their peers when the time for them to complete the test became shorter. This was perhaps because the students worried that they did not have enough time to complete the test. Accordingly, the teacher's prevention of her students' cheating was also escalated. She started her prevention by knocking her ruler at her desk and reminded the students that they were not allowed to talk when taking a test. Then she knocked her ruler harder, approached the whispering students, and threatened to punish them by subtracting two marks of their test results. The escalation of the number of students talking to their peers suggested that the students were very concerned about their test results, while the escalation of the teacher's prevention of her students' cheating suggested that she tried to ensure that fairness and accuracy of test results were attained. The teacher's threat to subtract marks from cheating students' test results suggested that she perceived that marks were high-stakes for her students.

This sub-section indicates that the teachers tried to ensure fairness in testing. They tried to ensure that the students took the same test in the same conditions. They also tried to stop their students from cheating when their students took a test. However, some teachers used marks to reward and punish their students.

8.3.2. Marks as rewards and punishment.

Some teachers in the observed classrooms used marks as rewards and punishment. Specifically, extra marks were given to those students who volunteered to answer the teachers' questions or who could do difficult tasks. The teachers also used marks as punishment for not doing homework.

8.3.2.1. Using marks as extrinsic rewards for the students' voluntary participation.

The teachers used marks as extrinsic rewards for those students who voluntarily answered their questions. In the following example, a teacher in School B rewarded her students by adding extra marks to a current oral mark of the students who volunteered to answer her questions.

In the previous lesson, Task 3 of the reading lesson in Unit 12 was not completed. The teacher had asked the students to undertake Task 3 (p. 138) of *English 11* at home.

> The teacher told the students that she would add two extra marks to their current oral marks if they volunteered to write the answers to the three questions in Task 3. Three students raised their hands. The teacher asked them to write the answers on the blackboard. Two of the students had three correct answers. The other student made one spelling mistake (i.e., he wrote "athelets" instead of "athletes").
> The teacher added two marks to the current oral marks of the students who had three correct answers and one mark to the current oral mark of the student who made a spelling mistake.

> (TB1Ob.T)

In this case, TB1 used marks as rewards for the students' voluntary participation in a classroom activity rather than as measures of the students' learning outcomes. This indicated that marks for this test were not accurate measures of learning outcomes. In addition, fairness was not ensured because the teacher did not assign marks as measures of learning outcomes for ranking the students as required by policy-makers. Even when using marks as extrinsic rewards was considered to be acceptable, the teacher unfairly treated the student who made a spelling error (i.e., she gave this student one extra mark instead of two extra marks) because the teacher assessed reading comprehension rather than spelling. The teacher's use of marks to reward students who volunteered to participate in classroom assessment was an unfair practice because timid students may not participate voluntarily and they may never be rewarded.

In another classroom observation, a teacher in School A used marks to challenge her students to do a difficult task.

> The teacher asked the students to listen to the tape and decide whether the five statements in Task 1 on page 128 of *English 10* were true or false.
> The teacher played the tape, then she asked the students to tell her which statements in Task 1 were true and which ones were false. The students could not give the correct answers.
> The teacher played the tape again. After the second listening, the students gave the correct answers to the first three questions, but they did not reach agreement on the answers to questions 4 and 5. The teacher said, "I will give you 10 marks if you can translate the statements that help you answer questions 4 and 5."
> This challenge was not met because none of the students could translate the statements in the tape that contained the answers to questions 4 and 5.

> (TA2Ob.L)

The students had difficulty understanding the part of the tape that contained the information that helped them respond to questions 4 and 5. However, TA2 challenged her students to do an even more difficult task (i.e., translating the statements into Vietnamese) by using marks as extrinsic rewards. The teacher's use of marks in this way suggested that she did not use marks as accurate measures of learning outcomes. Rather, she used marks to challenge her students to do tasks that she considered to be difficult. Using marks for this purpose may encourage some students to take risk but it may be unfair to other students in the class, especially timid students.

8.3.2.2. Using marks to punish the students for not doing homework.

Unfairness was also manifested through some teachers' use of marks as punishment. In one observed lesson, a teacher in School A punished her students for not doing homework.

> After greeting the students, the teacher asked a student to bring his notebook to her. The student stood up scratching his head and said, "I am sorry. I did not do the homework." The teacher said that she was not happy. She wrote the first sentence in Exercise 3 on page 131 of *English 10* on the blackboard. The sentence read as follows:
> If it rains, I'll stay at home and watch TV.
> The teacher asked the student to make a question that elicited the information in the underlined part of the sentence. The student could not make such a question. The teacher asked two other students to stand up and make the question, but neither of them could make the question.
> The teacher started complaining that some students were lazy and did not try to learn English because English would not be one of the three subjects that they would take in the NUEE.
> Then she said to the whole class, "Stand up if you did not complete Exercise 3 as I had told you to do last time." Six students stood up, including the three students who could not make the question as the teacher had required. The teacher said that she would subtract one mark from these students' oral marks that she had in her score book.
>
> (TA2Ob.R)

TA2 punished her students for not doing homework by decreasing the oral marks that they had obtained from the previous oral tests. This suggested that marks were not measures of the students' learning outcomes. Rather, marks were considered to be extrinsic rewards for effort. The use of marks as rewards and punishment for the students' effort may motivate some students to try harder. However, it was unfair to other students because in these schools, marks were used to rank the students and the different rankings were assumed to be indications of the students' competence.

This section indicates that the teachers tried to provide their students with equitable conditions when they gave them a 15-minute test. In these 15-minute tests, some students cheated by talking to each other. The teachers tried to stop cheating by monitoring the students and using marks to threaten them. The teachers' provision of equitable conditions and prevention of cheating suggested that they valued fairness in testing. However, some teachers used marks as rewards for the students' voluntary participation in classroom assessment. Some teachers also used marks to

punish those students who did not do their homework. The teachers' use of marks as rewards and punishment indicated that marks were not accurate measures of learning outcomes.

8.4. Discussion

The following findings were obtained from the results of the analysis of the classroom observations. First, apart from occasional quizzes, the teachers in this study relied on the tasks and exercises in the textbooks for assessment purposes. This finding was different from the finding in a study that McMillan (2001) carried out in Virginia in the USA. Most of the Grade 6 to 12 teachers in McMillan's studies designed their own assessments, and some used published tests for classroom assessment. The teachers' reliance on the exercises and tasks in the textbooks may derive from their compliance with the MOET's and the DOET's regulations about how they should use the textbooks. The MOET and the DOET require teachers to complete all the exercises in the textbooks. This reliance may also be due to a lack of knowledge and skills in assessment as remarked on by the principals in the study. This may be the case because training in assessment was not provided to pre-service teachers and the professional development documents did not describe how in-service teachers should use assessment for teaching and learning improvement.

Second, the teachers in this present study mainly assessed lower-order thinking skills. Specifically, they focused on the students' comprehension and especially recall of factual information, vocabulary, and grammar rules. This finding was similar to a finding in Kahn's (2000) study. In Kahn's study, the teachers teaching Grade 10 English students at a large suburban high school in the Midwest of the US mainly assessed their students' recall. These findings support Brookhart's (2010) argument that most teachers mainly assess students' recall of factual information because it is easier for them to formulate questions that elicit students' recall. The teachers may have had difficulty formulating questions that elicited students' higher-order thinking skills because they had not been taught how to do so in their pre-service teacher education programmes or through professional development programmes.

Third, the EFL teachers in the present study provided their students with only two types of feedback, i.e. direct correct feedback and praise about the students as persons. While feedback can be about different objects (i.e., the task, the processes used to complete the task, self-regulation, and "self"), and they have different effects on learning (Hattie & Timperley, 2007), the teachers provided the students with only corrective feedback and praise about the students as persons. Direct corrective feedback is not helpful for students to improve their learning (Shute, 2008b), and feedback on the students as persons is the least useful feedback (Hattie & Timperley, 2007). Therefore, the feedback provided by the teachers in this study was not useful for improving

students' learning, as perceived by the principals and students in this study. Perhaps the teachers did not provide their students with other types of feedback because they lacked the knowledge and skills to provide these types of feedback due to inadequate training and professional development.

Fourth, the teachers tried to prevent cheating by providing the same conditions for all the students when they administered a formal test and by threatening to subtract marks from the final mark that the students obtained from the test. The way the teachers tried to prevent cheating indicated that fairness was understood as providing the same conditions. This understanding indicated that assessment of learning was valued. Fairness in assessment for learning should be understood as providing all students with ample opportunities to learn (Baharloo, 2013; Lantolf & Poehner, 2013).

Fifth, the teachers in the present study used marks as rewards and punishment. Using marks as rewards and punishment is considered to be an abuse of teachers' power (Bandaranayake, 2011), but teachers in other contexts have also been found to use marks as rewards and punishment. For example, the teachers in a study by Davis and Neitzel (2011) in south-eastern USA also used marks as rewards and punishment. The finding in Davis and Neitzel (2011) and the finding in this study suggest that teachers in different contexts use marks to control their students' learning. The teachers' use of marks as rewards and punishment suggested that assessment in these schools was high stakes to the students and was administered mainly for summative purposes.

Sixth, cheating was observed in the classrooms when the students took a test that was administered to generate marks. This practice was consistent with the students' and teachers' reports about the students' cheating in tests. Cheating may be common in these schools because test results were used to rank, promote, and retain students. In addition, the teachers and students perceived that test results had consequences for students, teachers, and parents.

The assessment practices observed in these EFL teachers' classrooms indicated that assessment in these schools was summative assessment in nature. Marks were considered as central to assessment in that the teachers used marks to reward and punish their students and the students tried to obtain high marks. Alternative assessment tasks that provided formative feedback were not given to the students for the purposes of enhancing their own learning.

The previous chapter discussed the students' perceptions of assessment in their contexts. This chapter analysed the EFL teachers' classroom assessment practices. The next chapter discusses the major findings of the study.

Chapter 9. Discussion and Conclusion

9.1. Introduction

The aim of this study was to investigate the contexts of assessment in Grade 10 to 12 EFL classrooms in two high schools in Vietnam. Specifically, the study answered the following questions:

1. What are the perceptions of the contexts of assessment held by the principals, the Grade 10 to 12 EFL teachers, and students in two high schools in Vietnam?

2. How do the Grade 10 to 12 EFL teachers implement assessment in their classrooms?

3. How do the principals, the Grade 10 to 12 EFL teachers, and students negotiate the multiple purposes of assessment in their schools and classrooms?

Five sets of data were collected and analysed to answer the research questions. These were policy documents, two interviews with two principals, six interviews with six EFL teachers, 12 focus group interviews with 12 groups of students, and 42 classroom observations. The data were analysed inductively using content analysis. Delandshere's (2001) conceptual framework was used to examine the interviews and classroom observations from technological, philosophical, sociological, and ethical dimensions.

This chapter discusses the findings from the results chapters. These findings are discussed in the context of other research on principals', teachers', and students' perceptions of assessment and the assessment practices of teachers. The chapter highlights the contribution of this study to the body of knowledge about assessment in high schools. This chapter also presents the limitations of the study and the implications for future research and practice.

9.2. Key Findings

The findings in this study are discussed in terms of the three research questions that guided the study.

9.2.1. Perceptions of the contexts of assessment.

The participants' perceptions of assessment in the schools in the study revealed that assessment was neither typical of a culture of assessment nor typical of a culture of testing. Specifically, the participants perceived that assessment helped support and improve teaching and learning and it was also used for accountability and managerial purposes. However, the perceptions showed that elements of a culture of testing were more visible than those of a culture of assessment.

The participants' perceptions revealed that assessment in these schools was marks-driven. Consequently, control, compliance, and distrust were prevalent in the principals' and teachers' perceptions.

9.2.1.1. Perceived purposes of assessment.

In general, the principals, teachers, and students held similar purposes of assessment. They perceived that assessment had four major purposes: professional purposes (i.e., providing information about teaching and learning and helping teachers and students modify and improve teaching and learning); preparation for the national examinations; accountability purposes (i.e., controlling teaching and learning and making teachers and students accountable); and managerial purposes (i.e., providing the school administrators with information about teaching and learning in order to make decisions about interventions, ranking students, and reporting to parents).

First, the principals perceived that assessment provided information about teaching and learning. They believed that this information helped teachers and students modify and improve their teaching and learning. A study of principals in Western Canada by Newton, et al. (2010) found that the principals perceived that large-scale assessment in their contexts was high-stakes, but at the same time they believed that it helped improve teaching and learning. Similarly, principals in Michigan in the USA (Shen et al., 2010) perceived that assessment provided information about teaching and learning, and this information helped teachers and students modify and improve teaching and learning. At the same time, they thought that assessment results were used to make schools and teachers accountable. Assessment in the schools in the current study was carried out to generate marks for ranking students and making decisions about the promotion and retention of students. However, the principals also believed that assessment supported and improved teaching and learning. This finding resonated with the findings in the studies by Newton, et al. (2010) and by Shen, et al. (2010), and consolidated the conclusion that principals believed that assessment supported and improved teaching and learning even in contexts where assessment was high-stakes.

Similar to the principals, the teachers in this present study perceived that assessment informed and improved teaching and learning. This finding was similar to findings in studies conducted in many other contexts. In Western countries, Grade 4 to 7 teachers in southern USA (Davis & Neitzel, 2011), Grade 7 to 12 teachers in Canada (Leighton et al., 2010), primary and secondary school teachers in the UK (Hargreaves, 2005), and primary school teachers in Barcelona in Spain (Remesal, 2007) perceived that assessment informed teachers and students about their teaching and learning. In Asia, junior secondary school teachers in Hong Kong held a similar perception (Cheng, 2006). The perception that assessment helped teachers and students modify and

improve teaching and learning was held by teachers in New Zealand (G. T. L. Brown & Harris, 2009), in New Zealand and Queensland, Australia (G. T. L. Brown & Lake, 2006), in the Netherlands (Segers & Tillema, 2011), and in Hong Kong (G. T. L. Brown, Kennedy, et al., 2009; Cheng, 2006). The finding that the teachers in the present study believed that assessment helped improve teaching and learning revealed that not only teachers in contexts where assessment was low-stakes, such as primary schools in New Zealand (G. T. L. Brown, 2004), but teachers in high-stakes contexts, such as Hong Kong (G. T. L. Brown, Kennedy, et al., 2009; Fai, 2009) and the schools in this present study, held the perception that assessment helped improve teaching and learning.

Many students in the present study perceived that assessment helped improve teaching and learning. They believed that assessment results informed the teachers about their teaching and the students' learning. They thought that this information helped the teachers modify their teaching and it also helped the teachers and the students improve learning. This finding was consistent with findings in other studies. In Moni, et al.'s (2002) study, Grade 8 students in Queensland perceived that assessment information helped them improve their learning. Similarly, New Zealand secondary students believed that assessment helped teachers and students improve teaching and learning (G. T. L. Brown, Irving, et al., 2009; Peterson & Irving, 2008). The finding in the present study and the findings from previous studies indicated that the perception that assessment enhanced teaching and learning was held by students in different contexts.

Second, the principals and a number of teachers perceived that classroom and school tests helped familiarise students with the content, format, and procedures of the national examinations. In this study, preparation for tests and examinations was regarded to be a purpose of assessment in schools when the principals perceived that they were high-stakes. Similarly, in a study conducted in Hong Kong principals thought that tests given to Grade 3, 6, and 9 students were intended to identify schools that needed assistance to raise their performance (Ngan et al., 2010). The Hong Kong principals in Ngan et al.'s (2010) study perceived that these tests were high-stakes, and they asked their teachers to familiarise their students with the content and format of the tests in order to help them obtain high marks. Again, similarly, elementary principals in Western Canada in Newton et al.'s (2010) study perceived that the results of large-scale tests in Canada were used to judge the effectiveness of administrative and instructional staff. Therefore, they believed that preparing students for these tests was necessary. In Vietnam, Grade 12 students took the GEDE before leaving school, and they were granted a certificate only if they passed the examination. Only students who passed the GEDE were allowed to take the NUEE to enter a university or college. The principals in this study perceived that passing these examinations was both the evidence and the target of

students' learning. The principals also believed that the rate of students who passed these examinations, especially the NUEE, influenced the schools' reputations. Therefore, it was not unusual for the principals to regard familiarising their students with these examinations to be an important part of assessment in their schools.

Similar to the principals, most teachers in this study perceived that they needed to prepare their students for the national examinations. Some teachers perceived that passing these examinations was their students' learning target and preparing students for these examinations was an important part of the schools' activities. Preparing students for tests and examinations was also perceived to be a purpose of assessment by teachers in other studies. In Kahn's (2000) study of the materials developed to assess their students, the English language teachers in a secondary school in the Midwest of the USA reported that they prepared their students for school and district tests because this preparation helped them maintain students' attention and cooperation. Grade 4 to 7 teachers in South-eastern US in Davis and Neitzel's (2011) study perceived that they needed to familiarise their students with the content and format of state-mandated tests because these tests were high-stakes and district-level and state-level administrators expected them to prepare their students for these tests. In a study of Hong Kong secondary teachers' perceptions of school-based assessment, Qian (2014) found that the English language teachers perceived that when results of school-based assessments were used as a component of the Hong Kong Certificate of Education Examination, a high-stakes examination in Hong Kong, the assessments became high-stakes. In addition, they perceived that their students treated school-based assessment as an additional examination rather than an opportunity for them to develop their language ability. Therefore, they believed that they needed to spend a lot of time preparing their students for these school-based assessments. The finding in this present study about the teachers' perception of preparing their students for the GEDE and the NUEE indicated that the teachers considered that the preparation of students for tests and examinations was an important part of assessment in their classrooms and schools. This was especially the case when they perceived that tests and examinations were high-stakes for their students, for themselves, and for the reputation of their schools.

Third, the participants in the present study perceived that assessment had accountability purposes. Specifically, the principals believed that assessment helped them control teaching and learning in their schools and thus made teachers and students accountable for their work. In a study of how elementary, middle, and secondary school principals in a school district in Michigan, the USA, most of the 16 principals participating in the interviews reported using student achievement data for making teachers accountable for their teaching (Shen et al., 2010). Similarly, the principals in Newton et al.'s (2010) study perceived that the data gathered from large-scale tests were used to

evaluate the effectiveness of administrative and instructional staff. In Hong Kong, where assessment was expected to be implemented to support teaching and learning and school-based assessment results were expected to be used for school self-evaluation, principals believed that policy-makers wanted to make schools accountable through school-based assessment results (Ngan et al., 2010). The finding in this study and those in other studies suggested that principals in various contexts perceived that assessment was used to make students, teachers, and school administrators accountable for their work.

The teachers in the present study perceived that assessment made both teachers and students accountable. They also reported that they used marks as rewards and punishment. The perception that assessment made schools, teachers, and students accountable was held by teachers in various contexts. In a study by G. T. L. Brown (2004), primary teachers in Auckland, New Zealand, perceived that assessment made schools and teachers accountable, but they did not believe that assessment made students accountable. The researcher argued that these primary teachers did not agree that assessment made students accountable because assessment in elementary schools in New Zealand was low-stakes and was intended for supporting and improving teaching and learning (G. T. L. Brown, 2004). In a study of perceptions of assessment held by teachers in 12 primary schools and 3 secondary schools in Hong Kong, Brown, Kennedy et al. (2009) found that the teachers believed that assessment made schools, teachers, and students accountable. These researchers argued that Hong Kong teachers agreed more with this perception than primary teachers in New Zealand and Queensland, Australia. This was because assessment in Hong Kong schools was high-stakes, while assessment in primary schools in New Zealand and Queensland at the time of the study was low-stakes. The finding about accountability purposes of assessment perceived by the teachers in the present study consolidated the findings in previous studies.

Similar to the principals and teachers, the students in the present study perceived that assessment made them accountable for their learning. They thought that their teachers used assessment to monitor and control their learning. Students in previous studies held similar perceptions. In Moni et al.'s (2002) study, Grade 8 students in Queensland perceived that their teachers used assessment to monitor their learning. In responding to questionnaires about students' perceptions of assessment in studies by Brown and Hirschfeld (2008) and Brown, Irving et al. (2009), secondary students in New Zealand believed that assessment made schools, teachers, and themselves accountable. Brown, Irving et al. (2009) argued that students believed that assessment made schools, teachers, and students accountable when they perceived that assessment had high-stakes consequences for them. This argument could be used as an explanation for the perception

held by the students in the present study who perceived that their assessment results were used to rank them and had consequences for themselves, their teachers, and parents.

The last finding about the perceptions of assessment purposes held by the participants in this study was that the principals, teachers, and students in the present study perceived that assessment had managerial purposes. The principals reported that assessment results helped them make decisions about facility investment, professional development for their teachers, and collaboration with students and parents in order to help students learn better. They also reported that assessment results were reported to parents to inform them about their children's learning. The principals in the study by Shen, et al. (2010) reported that they relied on student achievement data obtained from standardised tests and classroom assessments for making decisions. Specifically, these principals used student achievement to make decisions about school improvement, to identify students' weaknesses, to help teachers improve their teaching, and work with parents and the community. Similarly, the principals in the elementary schools in Newton et al.'s (2010) study believed that data collected from large-scale assessments helped them identify students' strengths and weaknesses and to place students in special programmes. These Canadian principals also reported that results of large-scale assessments informed their school-based planning and priority-setting. When assessment results were used as data for managerial decisions, they should be accurate measures of students' learning outcomes. However, in the schools under investigation, marks were perceived not to reflect students' learning outcomes. Therefore, the administrators in these schools relied on unreliable data for making decisions.

The teachers in the present study perceived that assessment meant generating marks for ranking students and reporting to parents, and some teachers perceived these to be the most important purposes of assessment in their contexts. Pryor and Lubisi (2002) reported that Grade 4 to 6 teachers in South Africa perceived that assessment was administered to obtain marks for reporting to parents, principals, and inspectors. Secondary Hong Kong teachers in Cheng's (2006) study held a similar perception. Additionally, these Hong Kong teachers believed that they would get support from parents when they reported students' learning outcomes to them. In contexts where teachers perceived that the main purpose of assessment was to generate marks for ranking students and reporting to parents, they may try to raise their students' marks as the teachers in the present study did.

The findings about the perceptions of assessment purposes held by the participants in this study were consistent with findings in previous studies. They add some knowledge about principals', teachers', and students' perceptions about purposes of assessment in different contexts. In addition, these findings indicate that the participants in this study perceived that the improvement

purposes of assessment were not incompatible with preparing students for the national examinations, making teachers and students accountable, and generating marks for decision-making, ranking, and reporting. The findings about the participants' perceptions of the improvement purposes of assessment suggested that a culture of assessment existed in these schools. However, their perceptions of examination preparation, accountability purposes, and managerial purposes of assessment indicated that a culture of testing was deeply rooted in their schools. In addition, their concerns about marks, accuracy of marks, objectivity, and fairness suggested that elements of a culture of testing dominated those of a culture of assessment in these schools. The perceptions about the purposes of assessment held by the participants in this study suggested that assessment of learning and assessment for learning should not be considered as two dichotomies. Rather, they may co-exist in contexts where assessment of assessment was still dominant while assessment for learning was promoted.

9.2.1.2. Marks-driven assessment.

The participants' perceptions of marks, the accuracy of marks, and the teachers' views about objectivity and fairness in testing indicated that assessment in these schools was marks-driven. Specifically, the teachers and students were concerned about marks. In addition, the principals, teachers, and students were concerned about the accuracy of marks as measures of learning outcomes, and thus they were concerned about tests and procedures for test administration. The principals' and the teachers' concerns about marks were associated with their concerns about objectivity and fairness in assessment. However, the findings about teachers' and students' perceptions of marks, accuracy, objectivity, and fairness indicated that these qualities were not achieved in their contexts.

9.2.1.2.1. Teachers' concerns about marks and their perceptions of other stakeholders' concerns about marks.

Many of the teachers expected their own students to obtain high marks from tests because they perceived that marks had high-stakes consequences for their students and themselves. Most teachers reported that marks were used to rank, promote, and retain their students. A few teachers perceived that their teaching was judged according to their students' marks. Moreover, the teachers perceived that marks were reported to parents. Some teachers believed that students' marks influenced parents' social status in their communities. These findings indicated that when teachers perceived that marks had consequences for their students, themselves, and parents, they expected their students to obtain high marks.

The majority of the teachers in the present study thought that the principals, teachers, and students were concerned about marks. They believed that the principals and parents expected students to obtain high marks. In addition, they perceived that in assessment their students were mainly concerned about marks. The teachers' belief that principals and parents expected students to obtain high marks and students cared a great deal about marks was another new finding in studies about teachers' perceptions of assessment. As such it suggests that in a context where assessment is administered to generate marks for high-stakes consequences, teachers believe that other stakeholders expect students to obtain high marks.

When teachers perceive that marks have important consequences, they may improve their students' marks by any means (Stobart, 2008). Therefore, some teachers in this study reported constructing 15-minute tests according to their students' competence and modifying test difficulty so that their students could obtain higher marks.

9.2.1.2.2. Students' concerns about marks and their perceptions of other stakeholders' concerns about marks.

Many students in the present study perceived that marks were not accurate measures of their learning outcomes because of cheating, random choice, test anxiety, and teachers' favour for those students who took private lessons in teachers' homes. Some students in a study conducted in Queensland, Australia, perceived that marks did not accurately reflect learning outcomes (Moni et al., 2002). These Australian students believed that marks were influenced by teachers' personal feelings about individual students (Moni et al., 2002), while the students in the present study attributed the inaccuracy of marks to students' cheating, random choice, test anxiety, and teachers' favour for those students who took private lessons in teachers' homes.

The majority of the students in the present study admitted that they were very concerned about marks. A number of students reported that they tried to obtain high marks by any means. Many students perceived that their marks had consequences for themselves, their teachers, and their parents. Specifically, most students thought that their learning was judged solely by marks. Additionally, they perceived that marks were used to rank them, and the ranking shaped their image in other people's minds. They stated that their marks influenced their parents' behaviour towards them (i.e., rewarding or scolding). Some students also thought that marks were part of the competition among students, thus they influenced how students judged each other. This finding was similar to a finding in Smith and Gorard's (2005) study. Some students in Smith and Gorard's study reported comparing their marks with peers' marks.

The students' concerns about marks and their perceptions of the consequences created by marks were new findings in studies about students' perceptions of assessment. When the students were mainly concerned about marks and they perceived that their marks had consequences not only for themselves but also for their teachers and parents, cheating for high marks was very likely because they tried to obtain high marks in order to avoid or mitigate these consequences.

Most students in the present study expressed preference for multiple-choice tests, even though they believed that multiple-choice tests could not assess their learning as accurately as teacher observations, peer observations, and direct teacher-student interactions. Some students preferred multiple-choice tests because they perceived that it was easier for them to complete these tests by asking their peers for answers or guessing and ticking answers randomly. Many students did not like open-ended tests and oral tests because they were not confident that their English was good enough for these types of tests. These findings were similar to the findings in Yildirim's (2004) study. The students in Yildirim's (2004) study thought that short-answer tests did not accurately measure their performance. Similar to the students in the present study, these Turkish students liked multiple-choice tests. However, the Turkish students liked the multiple-choice format because it was used in university entrance examinations in Turkey. Like the students in the present study, the Turkish students thought that they could complete multiple-choice tests by guessing answers. The Turkish students did not like open-ended tests because they thought that unfairness may occur when teachers marked their test papers without fixed criteria. Moreover, these students thought that it was more difficult to answer open-ended questions and they felt uncomfortable when orally responding to teachers' questions in front of their peers. The students' mention of cheating as a reason for their preference of multiple-choice tests suggested that they were more concerned about completing tests and obtaining high marks than demonstrating their competence and learning through assessment.

Many students in the present study perceived that their teachers and parents expected them to obtain high marks. A number of students thought that their marks influenced how their teacher's teaching was judged. Several students perceived that their marks made their parents proud of them and even provided their parents with a social status within their home community. This finding challenged a finding in a study by Smith and Gorard (2005), which found that students were concerned about marks, even though they did not perceive that marks had high-stakes consequences.

9.2.1.2.3. Principals', teachers', and students' concerns about accuracy, objectivity, and fairness.

The principals in this study expected that marks were accurate measures of students' learning outcomes. They believed that 45-minute tests and end-of-term tests should have a broad coverage of knowledge. In addition, they thought that tests needed to be based on the MOET's standards of knowledge and skills so that test results accurately reflected students' learning outcomes. Indeed, no single test can cover all domains that need to be assessed because each test can only assess a sample of knowledge in a domain (Harlen, 2007). Therefore, the expectation that tests should have a broad coverage of knowledge was too ambitious. In addition, test results are influenced by a multitude of factors (Estaji, 2011; Must & Must, 2013; Torff & Sessions, 2009). Therefore, the principals' expectation that test results were accurate measures of the students' learning outcomes indicated that the principals cared much about measuring learning outcomes against standards and were concerned about marks. While the principals in the present study were concerned about standards in assessment, the American principals in a study by Militello, Bass, Jackson, and Wang (2013) reported that they did not care about aligning their assessments with state standards. Therefore, this finding extended the understanding about principals' perceptions assessment. Like the principals, a number of teachers in the present study valued the accuracy of marks. This concern was associated with their perceptions of the consequences of marks for their students, themselves, and parents.

Unlike the principals and the teachers, the students in the study did not express concerns about the accuracy of marks. Instead, most of the students perceived that marks were not accurate measures of their learning outcomes, even though a few students believed that marks were accurate measures of learning outcomes of good students only. Most of the students believed that marks obtained from multiple-choice tests did not reflect their learning outcomes because some students cheated and they could have guessed and ticked answers randomly. Many of them believed that anxiety caused by the pressure to do well in tests influenced their test results. Students in previous studies (Moni et al., 2002; Yildirim, 2004) also believed that marks were not accurate measures of students' learning outcomes.

The principals in the present study were concerned about their teachers' objectivity and fairness in assessment. These concerns were manifested in the procedures that they mandated for test administration. They mandated that the teachers use the schools' tests for 45-minute tests and end-of-term tests, even though the principals perceived that these mandates deprived the teachers of the opportunities to use assessment for supporting and improving teaching and learning. The sacrifice of the teachers' autonomy to use assessment to improve teaching and learning suggested

that in contexts where assessment results were high-stakes, the desire for objectivity and fairness subordinated the use of assessment for formative purposes.

Similar to the principals, some of the teachers valued objectivity and fairness in assessment. They reported supporting the schools' procedures intended for ensuring that these qualities were achieved. There have been many studies about teachers' perceptions and practices of assessment. However, teachers in previous studies did not express concerns about objectivity and fairness in assessment. Therefore, this is a new finding of the present study.

Although accuracy, objectivity, and fairness concerned the principals and many teachers, these qualities were not achieved in assessment in these schools. The principals thought that their teachers did not construct tests, especially 15-minute tests, according to the MOET's standards. They believed that the teachers constructed tests according to their own expectations of the students and according to their beliefs about the students' competence. The principals' perceptions of the teachers' test construction indicated that the marks obtained from the teacher-constructed tests were not accurate measures of students' learning outcomes against the MOET's standards.

The teachers' accounts about the way they constructed 15-minute tests further indicated that accuracy, objectivity, and fairness were not achieved in these contexts. Most teachers admitted constructing 15-minute tests according to their own expectations of the students' abilities and the students' marks from previous tests. Some teachers reported adjusting the difficulty of their 15-minute tests so that their own students could obtain higher marks. A number of teachers reported using marks to encourage their students to participate in classroom activities and to punish them for not doing homework. When different teachers based their tests on their own standards rather than on agreed standards, the difference in marks obtained by different students was not likely to be a reflection of the students' abilities. Unfairness resulted from the way the teachers constructed tests and the way they used marks for reward and punishment. Therefore, the way the teachers constructed tests was problematic. This often resulted in control from policy-makers and principals.

9.2.1.3. Control, compliance, and distrust.

Evidence of control, compliance, and distrust in assessment in these schools was pervasive in the principals' and the EFL teachers' perceptions. The principals valued their teachers' compliance with the MOET's, the DOET's, and the schools' regulations and mandatory procedures for assessment. Additionally, the principals equated teachers' knowledge and skills in assessment with their compliance with the regulations. Therefore, the teachers in this study were treated as technicians who implemented regulations issued by the MOET, the DOET, and the school administrators.

The principals' expectation of their teachers' compliance resulted from their distrust of the teachers' professional competence, professional commitment, and fairness in assessment. They perceived that the teachers had restricted knowledge and skills in constructing multiple-choice tests and using assessment to identify students' knowledge gaps. They also thought that the teachers' command of English was not very good. In addition, the principals believed that the teachers did not provide the students with quality feedback because they were not committed to their job. Further, the principals distrusted some teachers in constructing tests and marking students' test papers. They thought the teachers were being unfair and attributed this unfairness to some teachers' offering private lessons in their homes. Researchers have argued that beliefs in each other's competence, responsibility, dependability, and reliability are necessary for trusting relationships (Moye, Henkin, & Egley, 2005). However, the principals in the present study did not regard their teachers to be fully competent and responsible. Principals' distrust of teachers' knowledge and skills in assessment has been found in other studies such as that of Yildirim (2004). However, the finding about the principals' distrust of the teachers' professional commitment and fairness was unique to this study.

Most of the teachers perceived that the school administrators controlled their assessment. The teachers thought that they had some autonomy in implementing oral tests and 15-minute tests, but they had to follow the format and timing mandated by the school administrators. In addition, they perceived that the school administrators controlled the construction, administration, and marking of the 45-minute and end-of-term tests. This finding was consistent with a study by Yildirim (2004). The teachers in Yildirim's study reported that their principals controlled the format of their tests and the way they implemented assessment in their classrooms. However, the teachers in Kahn's (2000) study were allowed to design assessments and use tests from different sources.

The teachers reported that they complied with the regulations and procedures. This was because they perceived that compliance was a social norm that they should follow and it ensured security in their job. The finding about the compliance of the teachers in the present study consolidated a finding in Yildirim's (2004) study. The teachers in Yildirim's study complied with their principals' expectations around assessment.

Some teachers distrusted other teachers' fairness in testing. They believed that other teachers gave their students easier 15-minute tests. This distrust was understandable because many teachers perceived that marks were used to judge their teaching and had high-stakes consequences for their students.

The students also expressed distrust of their teachers' fairness in assessment. A number of students thought that in test construction, some teachers may favour the students who took private extra lessons in the teachers' homes. Some students believed that their teachers may give an

advantage to their own students when they marked their students' test papers, and this was especially the case for those who took private lessons with these teachers, The secondary students in Moni et al.'s (2002) study believed that marks may be influenced by teachers' personal feelings about individual students. However, the students in the present study distrusted their teachers' fairness mainly because some teachers gave private lessons in their homes.

Two important conclusions could be drawn from the findings in the present study related to the first research question. First, marks strongly influenced the participants' perceptions of assessment in their contexts. Assessment in the schools in this study was administered mainly to generate marks. Most students and teachers perceived marks to have high-stakes consequences for students, teachers, and parents. Such perceived consequences of marks made teachers and students concerned about marks. Such consequences also made the principals, teachers, and students concerned about the accuracy of marks and teachers' objectivity and fairness in testing. The perceived consequences of marks, however, resulted in cheating. In addition, some teachers manipulated their tests and used marks as rewards and punishment. Second, the principals' expectation of their teachers' compliance due to the principals' distrust of the teachers' competence, commitment, and fairness restricted the teachers' autonomy in assessment. In these schools, assessment for learning was a secondary purpose because the principals emphasised accuracy, objectivity, and fairness and expected their teachers to comply with the regulations and procedures in assessment. In addition, for job security, the teachers strictly complied with these regulations and procedures, even though they perceived that these regulations and procedures restricted their autonomy in assessing their students for learning improvement. Some of the students expected feedback that would assist their learning improvement, but most were more concerned about their marks.

9.2.2. The EFL teachers' assessment practices.

The classroom observations revealed that the EFL teachers in the study relied on the exercises in the textbooks for classroom assessment. The primary thinking skill that was assessed in the classrooms was recall of factual information, vocabulary, and grammar rules. The teachers' feedback included direct corrective feedback and praise. In addition, the teachers used marks as rewards, punishment, and threats.

9.2.2.1. Reliance on textbooks.

Apart from occasional quizzes used for assessment purposes, the EFL teachers in this study relied on the exercises in the textbooks when they implemented assessment in the observed classrooms. This assessment practice was different from the assessment practice in EFL classrooms

in the study of primary schools in Vietnam (Pham, 2013). Pham (2013) found that EFL teachers in three primary schools in Hanoi used various assessment methods in their classrooms. However, it should be noted that assessment in these primary schools was low stakes, and the teachers had considerable autonomy in deciding the content, timings, and assessment methods (Pham, 2013).

Researchers have found that teachers' assessment practices are influenced by a variety of factors. First, teachers' assessment practices are influenced by the assessment methods that are prioritised in the contexts in which they work (Cheng, 2006; Kahn, 2000). Second, teachers' knowledge and skills about assessment, especially their knowledge and skills about formative assessment, affect their assessment practices (Inbar-Lourie & Donitsa-Schmidt, 2009). Third, teachers' assessment practices are influenced by the subjects that they teach (McMillan, 2001). The high-stakes tests and examinations used in specific settings also exert a strong influence on teachers' assessment practices (Kahn, 2000; McMillan, 2001). Teachers' perceptions of the value and effectiveness of a certain method of assessment have a powerful impact on their assessment practices (Gattullo, 2000). Last but not least, school administrators' expectations influence teachers' assessment practices in their classrooms (Yildirim, 2004). The teachers' reliance on the exercises and tasks in the textbooks may derive from their compliance with the MOET's and the DOET's regulations about how they should use the textbooks. The MOET and the DOET require teachers to complete all the exercises in the textbooks. This reliance may also be due to a lack of knowledge and skills in assessment as remarked on by the principals in the study. This may be the case because training in assessment was not provided to pre-service teachers and the professional development documents did not describe how in-service teachers should use assessment for teaching and learning improvement.

9.2.2.2. Assessing lower-order thinking skills.

The EFL teachers in the present study focused on assessing students' lower-level thinking skills. Specifically, they assessed students' recall of factual information, vocabulary, and English grammar rules. Teachers in previous studies have also focused on students' memory and recall (Kahn, 2000; Pham, 2013) and comprehension (Gattullo, 2000; McMillan, 2001). Brookhart (2010) has argued that teachers' focus on assessing students' lower-order thinking skills, especially recall of factual information, is a "prevalent" (p. 2) phenomenon because questions that elicit this thinking skill are the easiest for teachers to ask. This argument may be used to explain why the teachers in the present study mainly assessed their students' recall of facts, vocabulary, and grammar rules. These teachers' focus on students' recall may also have been influenced by the exercises in the textbooks, which mainly assessed lower-order thinking skills. It may also have been influenced by

the thinking skills officially assessed in the schools' tests. The skills that had to be assessed were knowledge, comprehension, and application of knowledge (see Chapter 5). Another reason was that the teachers had to prepare their students for high-stakes tests and examinations, which often focus on lower-order thinking skills (Harlen, 2007). Anderson and Krathwohl (2001) argue that teaching/learning for retention and teaching/learning for transfer are "[t]wo of the most important educational goals" (p. 63). By retention they mean students can remember what they have learned, while transfer means students can use what they have learned in new situations and this is associated with "meaningful learning" (Anderson & Krathwohl, 2001, p. 63). In other words, teaching/learning for retention focuses on lower-order thinking skills, while teaching/learning for transfer focuses on higher-order thinking skills. These authors argue that teaching/learning for transfer is more desirable than teaching/learning for retention. The focus of assessment on the students' recall of factual information, vocabulary, and grammar rules indicated that the EFL teachers in this study focused on retention rather than on transfer.

9.2.2.3. *Providing direct corrective feedback and praise.*

Feedback has a powerful influence on student learning and achievement (Hattie & Timperley, 2007), and different students benefit from different types of feedback (Frey & Fisher, 2011; Hattie & Timperley, 2007; Shute, 2008b). However, the vast majority of the feedback provided by the teachers in the present study comprised direct corrective feedback. That is, the teachers indicated students' errors and provided correct answers to their students. This type of feedback was also the most common feedback given by the Hong Kong secondary English language teachers in Lee's (2003) study. Researchers have shown that direct corrective feedback is not as effective as feedback that shows students how to do a task (Hattie & Timperley, 2007; Shute, 2008b). Identifying students' errors and providing correct answers may be more time-saving than asking students to explain how they approach a question and how they come to an answer. However, direct corrective feedback does not help students see why they made an error and what they need to do to improve their learning. Therefore, many students in this study expressed dissatisfaction with their teachers' feedback. They expected their teachers to give them feedback that helped them know why they made an error and how to correct it. Instead, the teachers provided them with marks, ticks, crosses, and correct answers.

In addition to direct corrective feedback, the students in the observed classrooms were praised as individuals. Black and Wiliam (1998a) have argued that praise is not effective for learning per se, although it may positively influence students' attitudes towards learning. Other researchers have found that praise, together with rewards and punishment, is the least effective

feedback for learning improvement (Hattie & Timperley, 2007). Perhaps the teachers believed that praise made their students confident, just as they thought that giving the students pass marks would make them confident and motivated to learn. The teachers' use of praise as feedback indicated that they valued students' performance goals rather than learning goals. When valuing students' performance goals, teachers encourage their students to learn in order to receive praise, positive evaluation, and rewards from other people (Lens, 2001; Shute, 2008b). In contrast, if they value students' learning goals, teachers encourage their students to learn knowledge and skills and master them in new situations rather than to learn for rewards and praise (Lens, 2001; Shute, 2008b). The negative side of using praise as feedback is that students often tried to avoid difficult tasks and challenges because they were afraid of failure (Shute, 2008b). In addition, students' learning is motivated and controlled by extrinsic rewards and praise rather than the desire to construct and master knowledge and skills (Lens, 2001; Shute, 2008b). In short, the feedback that the EFL teachers gave their students in the observed classrooms indicated that they did not have the knowledge and skills in using feedback of various types in order to help the students improve their learning. This practice, according to the principals, was also because the teachers were not professionally committed.

9.2.2.4. Using marks as rewards, punishment, and threats.

In the observed classrooms, the teachers used marks to reward those students who voluntarily participated in classroom activities and to punish those students who did not do their homework. Using marks as rewards and punishment indicated that marks were measures of the students' effort rather than measures of their learning outcomes. However, when the teachers implemented 15-minute tests to generate marks, they provided all the students in the classroom with the same testing conditions in order to ensure fairness. In addition, they prevented and stopped their students who asked peers for answers. Some teachers threatened to deduct marks from the marks that the students obtained from the test if they cheated when taking a test. These observed activities suggested that some teachers perceived that their students were concerned about marks. Therefore, they used marks as a tool to exercise their power in controlling students' behaviours.

The classroom observations indicated that even though the elements of a culture of assessment existed in these schools (i.e., some assessments were embedded in instruction, and feedback was given when learning was occurring), the elements of a culture of testing were prominent. Specifically, marks were used to reward and punish students. In addition, tests were implemented to generate marks. Tests were used because the teachers thought that they needed to provide their students with equitable conditions and prevent cheating. However, this was not the

case because some students reported that students engaged in cheating. In addition, cheating occurred in the classroom tests when some students asked their peers for answers.

9.2.3. Negotiating the multiple purposes of assessment.

The findings of the study indicated that there were contradictions between what the principals, the EFL teachers, and the students in this study perceived they were expected to do, what they valued, what they reported doing, and what they did in assessment. Although these stakeholders perceived that assessment had multiple purposes, they emphasised the purposes of assessment that were typical of a culture of testing.

9.2.3.1. How the principals negotiated the purposes of assessment.

The principals perceived that assessment had multiple purposes but they prioritised the purposes and the elements that were typical of a culture of testing. On the one hand, they thought that assessment provided information about the effectiveness of the teachers' teaching and the students' learning outcomes. They believed that this information was necessary for the modification of the teachers' teaching and the improvement the students' learning. This was consistent with the purposes of assessment in a culture of assessment in that assessment was an integral part of instruction for the purpose of improving teaching and learning. On the other hand, they restricted the teachers' autonomy in implementing assessment for these purposes. The principals perceived that this restriction deprived the teachers of the opportunities to use assessment for gathering information about the students' learning so that they could modify their teaching and provide timely feedback for the students to improve their learning. Specifically, the principals did not allow the teachers to construct 45-minute tests and end-of-term tests for their own students, even though they believed that allowing the teachers to do so would help them identify the students' knowledge gaps and provide the students with timely feedback.

The principals' restriction of the teachers' autonomy resulted from the principals' emphasis on the accuracy of marks and their distrust of teachers' objectivity and fairness in assessment. The principals expected that tests accurately measure students' learning outcomes. In addition, they wanted their teachers to be objective and fair to all students. However, they thought that some teachers may favour some of their students if they were allowed to construct 45-minute tests and end-of-term tests. The principals mandated that the students in the same grade in the school take the same test at the same time. These tests were chosen from the schools' test banks. Moreover, the principals required that students' test papers be marked anonymously.

Perhaps the principals' emphasis on accuracy, objectivity, and fairness was a response to the MOET's call for these qualities in testing rather than the principals' genuine expectations. Indeed,

they expected that their students obtain high marks. Specifically, they expected their teachers to ensure that marks obtained by students in different classes were similar, even though the students had different abilities. Some teachers perceived that their principals expected the students to obtain high marks. In addition, they reported that the principals set a percentage of students with pass marks. The teachers perceived this target to be related to the principals' expectation of the teachers' manipulation of tests and marks in order to raise the students' test results.

In short, the principals perceived that assessment served various purposes, but they regarded it mainly as a tool to generate marks instead of a means to support and improve teaching and learning. Their expectation of the students' marks indicated that the emphasis on accuracy, objectivity, and fairness was a way to respond to the MOET's expectation rather than their own expectation.

9.2.3.2. *How the EFL teachers negotiated the purposes of assessment.*
Similar to the principals, the EFL teachers perceived that assessment had multiple purposes. One of the purposes that the teachers perceived was that assessment provided information about their teaching and students' learning. They believed that this information was helpful for them to modify their teaching and improve the students' learning. They also thought that marks needed to accurately reflect the students' learning outcomes. In addition, they expressed agreement with the schools' procedures intended for ensuring objectivity and fairness in assessment (e.g., using the schools' tests for 45-minute tests and end-of-term tests, giving all the students in the same grade the same 45-minute test or end-of-term test at the same time).

However, they reported that they manipulated tests to raise their students' marks. They admitted that they wanted their students to obtain high marks from tests because they perceived that their students' marks had consequences for the students, parents, and themselves. With this perception of the high-stakes consequences of their students' marks, the teachers reported that they manipulated 15-minute tests in order to raise their students' marks. The teachers in the study reported that they constructed 15-minute tests according to their students' competence. They also reported reducing the difficulty of their tests if their students obtained low marks from previous tests because they did not want their students to be disadvantaged compared with their colleagues' students. In addition, the teachers perceived that their principals expected them to give their students high marks. Some teachers acknowledged that marks were not accurate measures of students' learning outcomes because they used marks as rewards for the students' voluntary participation in classroom activities rather than as measures of the students' quality of work. The teachers' manipulation of tests were understandable because the high-stakes consequences attached

to students' marks drove the teachers to improve their students' marks in ways that may be inconsistent with what they really wanted to do (Madaus, Russell, & Higgins, 2009; Stobart, 2008).

In short, the teachers reported that they strictly complied with the regulations and procedures for assessment. These were mainly intended for ensuring that accuracy, objectivity, and fairness were achieved. However, the teachers implemented assessment mainly to generate marks for ranking students and reporting to parents. They had their own way to raise their own students' marks because they perceived that their students' marks had consequences for their students and for themselves.

9.2.3.3. How the students negotiated the purposes of assessment.

Like students in other contexts (G. T. L. Brown & Hirschfeld, 2007; G. T. L. Brown, Irving, et al., 2009), the students in the present study perceived that assessment had various purposes. However, they were mainly concerned about marks. Perceiving that marks had consequences for themselves, their teachers, and their parents, they were more concerned about completing tests and obtaining high marks than about demonstrating what they knew and could do in order to receive feedback to improve their learning. Specifically, they preferred multiple-choice tests because they perceived that it was easier to complete these tests. When taking multiple-choice tests, they had more opportunities to ask their peers for answers and they may have more chances to respond to given answers correctly and they may obtain high marks. The students' perceptions of the high-stakes consequences attached to marks and the pressure to do well in tests to obtain high marks drove some of them to cheat when they took tests. These reasons for the students' cheating in tests were consistent with researchers' argument that the likelihood of cheating was in proportion to the consequences attached to test results (Estaji, 2011; Madaus et al., 2009).

In summary, the principals, teachers, and students in this study were concerned about marks, and they paid more attention to raising test results than to using assessment for supporting and improving teaching and learning. Therefore, the principals, teachers, and students in this study emphasised the accountability and managerial purposes rather than using assessment for formative purposes.

9.3. Limitations of the Study

Some limitations of this study are acknowledged. First, due to the participant teachers' busy schedules, interviews following up the classroom observations could not be conducted. Such interviews might have been useful for the researcher to investigate the EFL teachers' articulation of their reasons why particular practices were followed, for example, why they relied on the exercises in the textbooks rather than designing their own assessments or using tests from other sources, why

they focused on lower-order thinking skills, and why they only gave direct corrective feedback and praise.

Second, this study was conducted in only two schools in one city in Central Vietnam, and the sample of participants was small. Therefore, the findings in this study cannot be generalised to other schools in Vietnam. The findings related to the assessment practices in the observed EFL classrooms in this study cannot be generalised to assessment practices in the classrooms of other subjects.

Third, as a study for a PhD thesis, there was no funding for inter-reliability checks to be undertaken and therefore these reliability checks were not carried out to make sure that the classroom observations truly reflected what the EFL teachers did in their classrooms.

Fourth, the percentage of the observed lessons was small (i.e., 15%) compared with the total lessons that the teachers taught in one school year. Hence the observed lessons may not have comprehensively reflected the teachers' classroom assessment. Intra-reliability checks were also not carried out. This is a limitation of the study.

9.4. Implications for Future Research

This study investigated the contexts of assessment in EFL classrooms in two high schools in Vietnam by looking at the principals', teachers', and students' perceptions of assessment; the EFL teachers' assessment practices; and the ways these stakeholders negotiated the multiple purposes of assessment in their schools. Based on this study, future research could explore the perceptions held by teachers of other subjects and their assessment practices in schools in Vietnam.

Although policy-makers have promoted assessment for learning at the school level in Vietnam, the professional development documents did not provide teachers with knowledge and skills in assessment for learning and the teachers in the present study had difficulty administering assessment for formative purposes in their classrooms. Future studies could focus on the challenges that EFL teachers in Vietnam have in implementing classroom assessment and what kind of support they need for implementing assessment for learning.

The analysis of the professional development documents indicated that assessment for learning was not part of professional development for EFL teachers in Vietnam. Future research could investigate teachers' professional training in assessment for learning and its impact on their perceptions and practices of assessment.

The principals in the present study tried to control their teachers' assessment practices but they did not mention how they helped their teachers improve knowledge and skills in assessment.

Future research could investigate what they do to help their teachers improve knowledge and skills in assessment, especially assessment for learning, and how they do it.

The findings in this study indicated that marks were the cause of students' cheating, teachers' test construction that did not follow the MOET's standards, and distrust among teachers and principals' and students' distrust of teachers. Future research could investigate whether or not marks positively influenced teachers' teaching and students' learning.

It was strange that none of the students in the present study mentioned preparation for the national examinations as a purpose of assessment in their schools, even though passing or failing these examinations had high-stakes consequences for them. Future research could investigate Vietnamese high school students' perceptions of these examinations and how they prepare for them.

9.5. Implications for Policy-Makers and Practitioners

The findings in this study suggested that if assessment for learning is to be promoted in EFL classrooms in high schools in Vietnam, policy-makers need to provide these teachers with knowledge and skills in how to implement classroom assessment and how to use data gathered from classroom assessment to support and improve teaching and learning.

Policy-makers may also need to reconsider how fairness is understood in Vietnamese high schools. This study found that principals and teachers thought that fairness was related solely to providing students with the same standardised testing conditions. While it might be argued that the use of standardized test conditions is necessary, policy-makers may wish to refer to a broader or alternative conceptualization of fairness. Such a viewpoint may mean that fairness is understood as providing students with ample opportunities for learning prior to assessing their knowledge and skills. In turn, this may maximise students' learning potential.

Principals have an important role in providing teachers with knowledge and skills in assessment, in deciding how assessment in their schools is implemented, and in using assessment data for different purposes. Therefore, with respect to implications for practitioners, principals themselves should develop their knowledge and skills in assessment, especially assessment for learning. Principals also should encourage teachers to implement assessment for the purpose of supporting and improving teaching and learning and allow teachers to have more autonomy in assessment.

Further, with reference to practitioners, the findings in this study indicated that these teachers lacked knowledge and skills in assessment, especially assessment for learning. Therefore, assessment literacy needs to be part of pre-service teacher training at colleges and universities in Vietnam. Teachers also need to improve their knowledge and skills in using multiple methods of

assessment, in providing quality feedback to students, and in using assessment data to modify their teaching and to help their students learn better. Collaboration among teachers in designing assessments and using assessment data may be key to developing teachers' knowledge and skills in assessment for learning. Teachers should involve their students in classroom assessment and help students assess themselves and their peers. The teachers in this study also relied on textbooks for their classroom assessment and their assessment mainly focused on assessing lower-order thinking skills. This suggests that EFL teachers need knowledge and skills in assessing students' higher-order thinking skills.

Students should be encouraged to be more responsible and active in assessment. They need support in using information from teacher feedback to improve their learning. In addition, they should be provided with guidance on how to assess themselves and their peers and get feedback from peers for learning rather than solely relying on teachers' assessment and feedback.

The students in this study reported that they wanted feedback that helped them improve their learning and they were not happy with the feedback that their teachers gave them. This suggests that quality feedback is welcomed by EFL students and they may benefit from this feedback. EFL teachers should develop knowledge and skills in assessment for learning in general and feedback in particular in order to help students improve their learning.

The students reported that they preferred multiple-choice tests because they thought that it was easier for them to complete them. However, they believed that other methods of assessment could measure their learning outcomes more accurately. This suggests that EFL students welcome various methods of assessment as long as they are not high-stakes. Therefore, EFL teachers should use various methods of assessment in their classrooms in order to help students improve their learning.

9.6. Concluding Remarks

This study found that in these two schools assessment was high-stakes and the principals, teachers, and students perceived it to have important consequences for students, teachers, and parents. When assessment was used and perceived this way, assessment results were the main concern for all stakeholders. The principals, teachers, and students were concerned about marks, thus they were concerned about the accuracy of marks and the teachers' objectivity and fairness. The concern about accuracy, objectivity, and fairness was so great that the teachers' autonomy in using assessment for supporting and improving teaching and learning was considered to be less important than ensuring that these qualities were achieved. However, the sacrifice of the teachers' autonomy in using assessment for supporting teaching and improving learning for accuracy,

objectivity, and fairness was not compensated. The teachers' test construction and use of marks as rewards and punishment and the students' cheating in testing made these qualities unachievable in these schools.

This study suggested that in contexts where assessment results were used for and perceived to have high-stakes consequences for students, teachers, and parents, it was difficult to develop a culture of assessment, even though policy-makers, principals, teachers, and students wanted it to be developed. As indicated by the findings of this study, assessment reform in Vietnam's high school education in general, and English instruction in particular, has not been very effective to date because major obstacles with respect to both perceptions and practices have not been tackled and removed. Reform of assessment in Vietnam has led to some changes in the methods of assessment rather than changes in the purposes of assessment. If the purposes of assessment are not changed and EFL teachers are not persuaded to change and provided with knowledge and skills in administering assessment for the purposes of enhancing teaching and learning, assessment of learning will continue to be prioritised and assessment for learning will be difficult to develop, no matter how much this is called for by policy-makers.

References

Agbatogun, A. O. (2014). Developing learners' second language communicative competence through active learning: Clickers or communicative approach? *Journal of Educational Technology & Society, 17*(2), 257-269.

Airasian, P. W. (1997). *Classroom assessment* (3rd ed.). New York: McGraw-Hill.

Alemi, M., Eslami, Z. R., & Rezanejad, A. (2014). Rating EFL learners' interlanguage pragmatic competence by non-native English speaking teachers. *Procedia - Social and behavioral Sciences, 98*, 171-174.

Allington, R. L., & McGill-Franzen, A. (1992). Unintended effects of educational reform in New York. *Educational Policy, 6*(4), 397-414.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Reseacrh Association.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives* (Complete ed.). New York: Longman.

Andrade, A. D. (2009). Interpretive research aiming at theory building: Adopting and adapting the case study design. *The Qualitative Report, 14*(1), 42-60.

Angrosino, M. V. (2007). *Naturalistic observation*. Walnut Creek, CA: Left Coast Press.

Angrosino, M. V., & Mays de Pérez, K. A. (2000). Rethinking observation: From method to context. In N. K. Denzin & Y. S. Lincoln (Eds.), *Collecting and interpreting qualitative materials* (2nd. ed., pp. 673-702). Thousand Oaks, CA: SAGE.

Anh, P. S. (2006). *Assessment in education.* Paper presented at the Assessment for Improving Secondary and High School Students' Activeness, Ho Chi Minh City University of Pedagogy, Vietnam.

Anh, V. T. P. (2006). *Assessment for learning: New trend in the world and lessons for Vietnam.* Paper presented at the Assessment for Improving Socondary and High School Students' Activeness, Ho Chi Minh City University of Pedagogy, Vietnam.

Assessment Reform Group. (2002). Assessment for learning: 10 principles., from http://www.assessment-reform-group.org.uk

Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., & Vleuten, C. P. M. V. (2007). Teachers' opinions on quality criteria for competency assessment programs. *Teaching & Teacher Education, 23*, 857-867.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing ueful language tests*. Oxford: OUP.

Bagarié, V. (2007). Defining communicative competence. *Metodika, 8*(1), 94-103.

Baharloo, A. (2013). Test fairness in traditional and dynamic assessment. *Theory and Practice in Language Studies, 3*(10), 1930-1938.

Bailey, C. A. (1996). *A guide to field research*. Thousand Oaks, CA: Pine Forge Press.

Bandaranayake, R. C. (2011). The ethics of student assessment. *Medical Teacher, 33*, 435-436.

Barjesteh, H., & Niknezhad, F. (2013). A pradigm shift toward a new philosophy of assessment: Dynamic assessment from a critical perspective. *Indian Journal of Fundamental and Applied Life Science, 3*(3), 526-535.

Baviskar, S. N., Hartle, R. T., & Whitney, T. (2009). Essential criteria to characterize constructivist teaching: Derived from a review of the literature and applied to five constructivist-teaching method articles. *International Journal of Science Education, 31*(4), 541-550.

Beck, D. E. (2000). Performance-based assessment: Using pre-established criteria and continuous feedback to enhance a student's ability to perform practice tasks. *Journal of Pharmacy Practice, 13*(5), 347-364.

Bell, J. (2005). *Doing your research project: A guide for first-time researchers in education, health and social sciences*. Berkshire: Open University Press.

Bennett, R. E., & Gitomer, D. H. (2009). Transforming K-12 assessment: Integrating accountability testing, formative assessment and professional support. In C. Wyatt-Smith & J. J. Cumming (Eds.), *Educational assessment for the 21st century: Connecting theory and practice* (pp. 43-61). New York: Springer.

Berry, R. (2011a). Assessment trends in Hong Kong: Seeking to establish formative assessment in an examination culture. *Assessment in Education: Principles, Policy & Practice, 18*(2), 199-211.

Berry, R. (2011b). Educational assessment in Mainland China, Hong Kong and Taiwan. In R. Berry & B. Adamson (Eds.), *Assessment reform in education: Policy and practice* (pp. 49-61). Dordrecht: Springer.

Biçak, B. (2013). Scale for test preparation and test taking strategies. *Educational Sciences: Theory & Practice, 13*(1), 279-289.

Binh Dinh DOET. (2009). *The English curriculum for high school students (Attached with 1716/SGDĐT-GDTrH: The curricula for secondary and high schools in the School year 2009-2010, issued by Binh Dinh Department of Education and Training on 09/9/2009)*. Binh Dinh DOET.

Binh Dinh DOET. (2010). *1898/SGDĐT-GDTrH: Guidelines on the implementation of end-of-term tests and ranking secondary and high school students in the School year 2010-2011*. Binh Dinh DOET.

Black, P. (1998). *Testing: Friend or foe? The theory and practice of assessment and testing*. London: Falmer Press.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Berkshire, UK: Open University Press.

Black, P., & Jones, J. (2006). Formative assessment and the learning and teaching of MFL: Sharing the language learning road map with the learners. *Language Learning Journal, 34*(1), 4-9.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74.

Black, P., & Wiliam, D. (1998b). Inside the blackbox: Raising standards through classroom assessment. *The Phi Delta Kappa, 80*(2), 139-148.

Black, P., & Wiliam, D. (2008). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*, 5-31.

Blackmore, J. (1988). *Assessment and accountability*. Geelong, Victoria: Deakin University Press.

Blanton, W. E., & Medina, A. (2009). Context in education. In E. F. Provenzo, J. P. Renaud, & A. B. Provenzo (Eds.), *Encyclopedia of the social and cultural foundations of education (Vol. 1-3)* (pp. 179-182). Thousand Oaks, CA: SAGE.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *A taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain* (B. S. Bloom Ed.). New York, NY: David McKay Company.

Bock, G. (2000). Difficulties in implementing communicative theory in Vietnam. *Teachers Edition, 2*, 24-30.

Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning and assessment. *Assessment & Evaluation in Higher Education, 24*(4), 413-426.

Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education: Principles, Policy & Practice, 8*(2), 153-169.

Brookhart, S. M. (2009). Assessment, gender and in/equity. In C. Wyatt-Smith & J. J. Cumming (Eds.), *Educational assessment in the 21st century: Connecting theory and practice* (pp. 119-136). New York: Springer.

Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classrooms*. Alexandria, VA: ASCD.

Brookhart, S. M., & Bronowicz, D. L. (2003). 'I don't like writing. It makes my fingers hurt': Students talk about their classroom assessments. *Assessment in Education: Principles, Policy & Practice, 10*(2), 221-242.

Brooks, J. G., & Brooks, M. G. (1999). *In search of understanding: The case for constructivist classrooms*. Alexandria, VA: Association for Supervision and Curriculum Development.

Brooks, M. G., & Brooks, J. G. (1999). The courage to be constructivist. *Educational Leadership, 57*(3), 18-24.

Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice, 11*(3), 301-318.

Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged version. *Psychological Reports, 99*, 166-170.

Brown, G. T. L., & Harris, L. R. (2009). Unintended consequences of using tests to improve learning: How improvement-oriented resources heighten conceptions of assessment as school accountability. *Journal of MultiDisciplinary Evaluation, 6*(12), 68-91.

Brown, G. T. L., & Hirschfeld, G. H. F. (2007). Students' conceptions of assessment and Mathematics: Self-regulation raises achievement. *Australian Journal of Educational & Developmental Psychology, 7*, 63-74.

Brown, G. T. L., & Hirschfeld, G. H. F. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy & Practice, 15*(1), 3-17.

Brown, G. T. L., Irving, S. E., Peterson, E. R., & Hirschfeld, G. H. F. (2009). Use of interactive-informal assessment practices: New Zealand secondary students' conceptions of assessment. *Learning and Instruction, 19*, 97-111.

Brown, G. T. L., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for student improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy & Practice, 16*(3), 347-363.

Brown, G. T. L., & Lake, R. (2006). *Queensland teachers' conceptions of teaching, learning, curriculum and assessment: Comparisons with New Zealand teachers*. Paper presented at the Annual Conference of the Australian Association for Research in Education (AARE), Adelaide, Australia.

Brown, G. T. L., Lake, R., & Matters, G. (2009). Assessment policy and practice effects on New Zealand and Queensland teachers' conceptions of teaching. *Journal of Education for Teaching, 35*(1), 61-75.

Brown, G. T. L., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching & Teacher Education, 21*(1), 210-220.

Brown, H. D. (2001). *Teaching by principles: An interactive approach to language pedagogy* (2nd ed.). White Plains, NY: Pearson Education.

Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. White Plains, New York: Pearson Education.

Brown, H. D. (2007). *Principles of language learning and teaching* (5th ed.). White Plains, NY: Peason Education.

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

Bulterman-Bos, J., Terwel, J., Verloop, N., & Wardekker, W. (2002). Observation in teaching: Toward a practice of objectivity. *Teachers College Record, 104*(6), 1069-1100.

Butler, Y. G. (2009). How do teachers observe and evaluate elementary school students' foreign language performance? A case study from South Korea. *TESOL Quarterly, 43*(3), 417-444.

Butler, Y. G., & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing, 27*(1), 5-31.

Caldwell, K., & Atwal, A. (2005). Non-participant observation: Using video tapes to collect data in nursing research. *Nurse Researcher, 13*(2), 42-54.

Camilli, G. (2013). Ongoing issues in test fairness. *Educational Research and Evaluation, 19*(2-3), 104-120.

Can, N. H. (2007). Teaching and learning of foreign languages in Vietnam: The current situation and some solutions. *Social Sciences Information Review, 1*(1), 43-52.

Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 2-27). London: OUP.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*, 1-47.

Canale, M., & Swain, M. (1981). A theoretical framework for communicative competence. In A. S. Palmer, P. J. M. Groot, & G. A. Trosper (Eds.), *The construct validation of tests of communicative competence* (pp. 31-36). Washington, DC: Teachers of English to Speakers of Other Languages.

Canh, L. V., & Barnard, R. (2009). Curricular innovation behind closed classroom doors: A Vietnamese case study (Vol. 24, pp. 20-33). Sydney: AMEP research Centre, Macquarie University.

Carr, M. (2008). Can assessment unlock and open the doors to resourcefulness and agency? In S. Swaffield (Ed.), *Unlocking assessment: Understanding for reflection and application* (pp. 36-54). New York: Routledge.

Chalhoub-Deville, M., & Deville, C. (2005). A look back at and forward to what language testers measure. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 815-832). Mahwah, NJ: Mahwah, NJ Lawrence Erlbaum Associates.

Chan, Y.-C. (2007). Elementary school EFL teachers' beliefs and practices of multiple assessments. *Reflections on English Language Teaching, 7*(1), 37-62.

Chappuis, J., Stiggins, R., Chappuis, S., & Arter, J. (2012). *Classroom assessment for student learning: Doing it right - Doing it well* (2nd. ed.). Boston: Pearson.

Chen, K. (2003). A constructivist approach to teaching: Implications in teaching computer networking. *Information Technology, Learning, and Performance Journal, 21*(2), 17-27.

Cheng, M. H. (2006). Junior secondary science teachers' understanding and practice of alternative assessment in Hong Kong: Implications for teacher professional development. *Canadian Journal of Science, Mathematics and Technology Education, 6*(3), 227-243.

Choi, I.-C. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing, 25*(1), 39-62.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education* (5th ed.). London: Routledge Falmer.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th. ed.). New York: Routledge.

Cohen, R. (1990). Reshaping the standards agenda: From an Australian's perspective of curriculum and assessment. In P. Broadfoot, R. Murphy, & H. Torrance (Eds.), *Changing educational assessment: International perspectives and trends* (pp. 32-51). London: Routledge.

Cooperstein, S. E., & Kocevar-Weidinger, E. (2004). Beyond active learning: A constructivist approach to learning. *Reference Services Review, 32*(2), 141-148.

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: SAGE.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approach* (3rd ed.). Los Angeles: SAGE.

Creswell, J. W. (2012). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (4th ed.). Boston: Pearson.

Crooks, T. (2011). Assessment for learning in the accountability era: New Zealand. *Studies In Educational Evaluation, 37*(2011), 71-77.

Cuong, L. A. (2006). *Assessment in active learning in secondary schools.* Paper presented at the Assessment for Improving Secondary and High School Students' Activeness, Ho Chi Minh City University of Pedagogy, Vietnam.

Curren, R. R. (1995). Coercion and the ethics of grading and testing. *Educational Theory, 45*(4), 425-441.

Dann, R. (2002). *Promoting assessment as learning: Improving the learning process*. New York: Routledge.

Davies, A. (1997). Introduction: The limits of ethics in language testing. *Language Testing, 14*(3), 235-241.

Davis, D. S., & Neitzel, C. (2011). A self-regulated learning perspective in middle grades classroom assessment. *The Journal of Educational Research, 104*(3), 202-215.

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly, 43*(3), 393-415.

Delandshere, G. (2001). Implicit theories, unexamined assumptions and the status quo of educational assessment. *Assessment in Education: Principles, Policy & Practice, 8*(2), 113-133.

Delandshere, G. (2002). Assessment as inquiry. *Teachers College Record, 104*(7), 1461-1484.

deMarrais, K. (2004). Qualitative interview studies: Learning through experience. In K. deMarrais & S. D. Lapan (Eds.), *Foundations for research: Methods of inquiry in education and the social sciences* (pp. 51-68). Mahwah, NJ: Erlbaum.

Denham, P. A. (1992). English in Vietnam. *World Englishes, 11*(1), 61-69.

Denzin, N. K., & Lincoln, Y. S. (2003). *Strategies of qualitative inquiry* (2nd ed.). Thousand Oaks, CA: SAGE.

Dicicco-Bloom, B., & Crabtree, B. F. (2006). The qualitative research interview. *Medical Education, 40*, 314-321.

Dixon, H., & Haigh, M. (2009). Changing mathematics teachers' conceptions of assessment and feedback. *Teacher Development, 13*(2), 173-186.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: OUP.

Downe-Wamboldt, B. (1992). Content analysis: Method, applications, and issues. *Health Care for Women International, 13*(3), 313-321.

Ellis, R. (2009a). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics, 19*(3), 221-246.

Ellis, R. (2009b). A typology of written corrective feedback types. *ELT Journal, 63*(2), 97-107.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing, 62*(1), 107-115.

Elwood, J. (2013). Educational assessment policy and practice: A matter of ethics. *Assessment in Education: Principles, Policy & Practice, 20*(2), 205-220.

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), 243-256.

Estaji, M. (2011). Ethics and validity stance in educational assessment. *English Language and Literature Studies, 1*(2), 89-99.

Fai, H. S. K. (2009). *Investigating missing conceptions of assessment: Qualitative studies with Hong Kong teachers*. Paper presented at the The European Association for Research on

Learning and Instruction (EARLI) 13th Biennial Conference, "Fostering Communities of Learning", Amsterdam, The Netherlands.

Fautley, M., & Savage, J. (2008). *Assessment for learning and teaching in secodary schools*. Exeter: Learning Matters.

Fisher, D., & Frey, N. (2011). Feed up, feedback, and feed forward. *Science and Children, 48*(9), 26-30.

Foster, P. (2006). Observational research. In R. Sapsford & V. Jupp (Eds.), *Data collection and analysis* (2nd ed., pp. 57-93). London: SAGE.

Fotos, S. (2005). Traditional and grammar translation methods for second language teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 653-670). Mahwar, NJ: Lawrence Erlbaum Associates.

Frey, N., & Fisher, D. (2011). *The formative assessment action plan: Practical steps to more successful teaching and learning*. Alexandria, VA: ASCD.

Gattullo, F. (2000). Formative assessment in ELT primary (elementary) classrooms: An Italian case study. *Language Testing, 17*(2), 278-288.

Gay, L. R., & Airasian, P. W. (2000). *Educational research: Competencies for analysis and application*. Englewood Cliffs: Prentice Hall.

Genesee, F., & Upshur, J. A. (1996). *Classroom-based evaluation in second language education*. New York: Cambridge University Press.

Gibson, W. J., & Brown, A. (2009). *Working with qualitative data*. London: SAGE.

Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer Press.

Gipps, C. (1999). Socio-cultural aspects of assessment. In P. D. Pearson & A. Iran-Nejad (Eds.), *Review of Research in Education* (Vol. 24, pp. 355-392). Washington, DC: American Educational Research Association.

Gipps, C., & Stobart, G. (2009). Fairness in assessment. In C. Wyatt-Smith & J. J. Cumming (Eds.), *Educational assessment in the 21st century: Connecting theory and practice* (pp. 105-118). NY: Springer.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report, 8*(4), 597-607.

Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: Concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today, 24*(2), 105-112.

Greene, J. C., Kreider, H., & Mayer, E. (2005). Combining qualitative and quantitative methods in social inquiry. In B. Somekh & C. Lewin (Eds.), *Research methods in the social sciences* (pp. 274-281). London: SAGE.

Griffin, P. (2009). Teachers' use of assessment data. In C. Wyatt-Smith & J. J. Cumming (Eds.), *Educational assessment in the 21st century: Connecting theory and practice* (pp. 183-208). New York: Springer.

Hammersley, M. (2008). Troubles with triangulation. In M. M. Bergman (Ed.), *Advances in mixed methods research* (pp. 22-36). Los Angeles: SAGE.

Hargreaves, E. (2005). Assessment for learning? Thinking outside the (black) box. *Cambridge Journal of Education, 35*(2), 213-224.

Harlen, W. (2007). *Assessment of learning*. London: SAGE.

Harlen, W. (2008). Trusting teachers' judgement. In S. Swaffield (Ed.), *Unlocking assessment: Understanding for reflection and application* (pp. 138-153). London: Routledge.

Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice, 4*(3), 365-379.

Harlen, W., & Winter, J. (2004). The development of assessment for learning: Learning from the case of science and mathematics. *Language Testing, 21*(3), 390-408.

Harris, L. R., & Brown, G. T. L. (2008). *New Zealand teachers' conceptions of the purpose of assessment: Phenomenographic analyses of teachers' thinking*. Paper presented at the The Australian Association for Research in Education (AARE) Annual Conference, Brisbane, Australia. www.aare.edu.au/08pap/har08257.pdf

Harris, L. R., & Brown, G. T. L. (2009). The complexity of teachers' conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice, 16*(3), 365-381.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Herbert, R. D., & Higgs, J. (2004). Complementary research paradigms. *Australia Journal of Physiotherapy, 50*, 63-64.

Hill, M. (2009). Ways of seeing: Using ethnography and Foucault's 'toolkit' to view assessment practices differently. *Qualitative Research, 9*(3), 309-330.

Hinkel, E. (2005). Introduction. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 631-634). Mahwar, NJ: Lawrence Erlbaum Associates.

Hittleman, D. R., & Simon, A. J. (2002). *Interpreting educational research: An introduction for consumers of research* (3rd ed.). New Jersey: Merrill Prentice Hall.

Hoffman, J. V., Assaf, L. C., & Paris, S. G. (2001). High-stakes testing in reading: Today in Texas, tomorrow? *The Reading Teacher, 54*(5), 482-492.

Horsburgh, D. (2003). Evaluation of qualitative research. *Journal of Clinical Nursing, 12*(2), 93-99.

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277-1288.

Huong, T. T. T. (2010). Insights from Vietnam. In R. Johnstone (Ed.), *Learning through English: Policies, challenges and prospects - Insights from East Asia*: British Council.

Huy, L. H. (2006). *Experimenting with improving assessment in Biology in Chuong Binh Le Private High School in An Giang Province.* Paper presented at the Assessment for Improving Secondary and High School Students' Activeness, Ho Chi Minh City University of Pedagogy, Vietnam.

Huyen, T. T. (2006). *The status quo of assessing high school students in An Giang Province.* Paper presented at the Assessment for Improving Secondary and High School Students' Activeness, Ho Chi Minh City University of Pedagogy, Vietnam.

Inbar-Lourie, O. (2008a). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing, 25*(3), 385-402.

Inbar-Lourie, O. (2008b). Language assessment culture. In N. H. Hornberger (Ed.), *Encyclopedia of language and education* (pp. 2418-2432). New York: Springer.

Inbar-Lourie, O., & Donitsa-Schmidt, S. (2009). Exploring classroom assessment practices: The case of teachers of English as a foreign language. *Assessment in Education: Principles, Policy & Practice, 16*(2), 185-204.

James, M. (2008). Assessment and learning. In S. Swaffield (Ed.), *Unlocking assessment: Understanding for reflection and application* (pp. 20-35). New York: Routledge.

Johnson, R., & Waterfield, J. (2004). Making words count: The value of qualitative research. *Physiotherapy Research International, 9*(3), 121-131.

Johnston, P., & Costello, P. (2005). Theory and research into practice: Principles for literacy assessment. *Reading Research Quarterly, 40*(2), 256-267.

Jones, B. D., & Egley, R. J. (2006). Looking through different lenses: Teachers' and administrators' views of accountability. *The Phi Delta Kappan, 87*(10), 767-771.

Kahn, E. A. (2000). A case study of assessment in a Grade 10 English course. *The Journal of Educational Research, 93*(5), 276-286.

Kane, M. (2010). Validity and fairness. *Language Testing, 27*(2), 177-182.

Kanjee, A. (2009). Enhancing teacher assessment practices in South African schools: Evaluation of the assessment resourse banks. *Education as Change, 13*(1), 73-89.

Kaufman, D. (2004). Constructivist issues in language learning and teaching. *Annual Review of Applied Linguistics, 24*, 303-319.

Kavanagh, B. (2012). The theory and practice of communicative language teaching in Japan. *Academic Research International, 2*(2), 730-738.

Kleinsasser, A. M. (1995). Assessment culture and national testing. *The Clearing House, 68*(4), 205-210.

Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice, 16*(3), 263-268.

Kopriva, R. J. (2008). *Improving testing for English language learners*. New York, NY: Routledge.

Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Cambridge, MA: Havard University Press.

Kozulin, A. (2002). Sociocultural theory and the mediated learning experience. *School Psychology International, 23*(1), 7-35.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: SAGE.

Kunnan, A. J. (2004). Test fairness. In M. Milanovic & C. Weir (Eds.), *European language testing in a global context*. Cambridge, UK: Cambridge University Press.

Kunnan, A. J. (2010). Test fairness and Toulmin's argument structure. *Language Testing, 27*(2), 183-189.

Lantolf, J. P., & Poehner, M. E. (2013). The unfairness of equal treatment: Objectivity in L2 testing and dynamic assessment. *Educational Research and Evaluation, 19*(2-3), 141-157.

Lee, I. (2003). L2 writing teachers' perspectives, practices and problems regarding error feedback. *Assessing Writing, 8*(3), 216-237.

Lee, I. (2007). Assessment for learning: Integrating assessment, teaching, and learning in the ESL/EFL writing classroom. *The Canadian Modern Language Review, 64*(1), 199-214.

Lee, I. (2008). Understanding teachers' written feedback practices in Hong Kong secondary classrooms. *Journal of Second Language Writing, 17*(2), 69-85.

Lee, I., & Coniam, D. (2013). Introducing assessment for learning for EFL writing in an assessment of learning examination-driven system in Hong Kong. *Journal of Second Language Writing, 22*, 34-50.

Leighton, J. P., Gokiert, R. J., Cor, M. K., & Heffernan, C. (2010). Teacher beliefs about the cognitive diagnostic information of classroom- versus large-scale tests: Implications for assessment literacy. *Assessment in Education: Principles, Policy & Practice, 17*(1), 7-21.

Lens, W. (2001). How to combine intrinsic task-motivation with the motivational effects of the instrumentality of present tasks for future goals. In A. Efklides, J. Kuhl, & R. M. Sorrentino (Eds.), *Trends and prospects in motivation research* (pp. 23-36). New York: Kluwer Academic Publishers.

Li, D. (2004). Trustworthiness of think-aloud protocols in the study of translation process. *International Journal of Applied Linguistics, 14*(3), 301-313.

Lietz, C. A., Langer, C. L., & Furman, R. (2006). Establishing trustworthiness in qualitative research in social work: Implications from a study regarding spirituality. *Qualitative Social Work, 5*(4), 441-458.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Thousand Oaks, CA: SAGE.

Lincoln, Y. S., & Guba, E. G. (2003). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *The landscape of qualitative research: Theories and issues* (pp. 253-291). Thousand Oaks, CA: SAGE.

Linn, R. (2000). Assessment and accountability. *Educational Researcher, 29*(2), 4-16.

Linn, R. L. (1986). Educational testing and assessment: Research needs and policy issues. *American Psychologist, 41*(10), 1153-1160.

Lutz, S. T., & Huitt, W. G. (2004). Connecting cognitive development and constructivism: Implications from theory for instruction and assessment. *Constructivism in the Human Sciences, 9*(1), 67-90.

Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing, 18*(4), 351-372.

MacBeth, D. (2001). On "reflexivity" in qualitative research: Two readings, and a third. *Qualitative Inquiry, 7*(1), 35-68.

Madaus, G., Russell, M., & Higgins, J. (2009). *The paradoxes of high stakes testing: How they affect students, their parents, teachers, principals, schools, and society*. Charlotte, NC: Information Age Publishing.

Maier, U. (2009). Towards state-mandated testing in Germany: How do teachers assess the pedagogical relevance of performance information? *Assessment in Education: Principles, Policy & Practice, 16*(2), 205-226.

Manly, J. J., Jacobs, D. M., Touradji, P., Small, S. A., & Stern, Y. (2002). Reading level attenuates differences in neuropsychological test performance between African American and White elders. *Journal of the International Neuropsychological Society, 8*, 341-348.

Mauther, N. S., & Doucet, A. (2003). Reflexive accounts and accounts of reflexivity in qualitative data analysis. *Sociology, 37*(3), 413-431.

McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice, 20*(1), 20-32.

McNamara, T., & Roever, C. (2007). *Language testing: The social dimension*. Maiden, MA: Blackwell.

McNamara, T., & Ryan, K. (2011). Fairness versus justice in language testing: The place of English literacy in the Australian Citizenship Test. *Language Assessment Quarterly, 8*(2), 161-178.

Merriam, S. B. (2009). *Qualitative research: A Guide to design and implementation*. San Francisco, CA: Jossey-Bass.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*(3), 241-256.

Middlehurst, R. (1997). Enhancing quality. In F. Coffield & B. Williamson (Eds.), *Repositioning higher education* (pp. 45-56). Buckingham: SRHE and Open University Press.

Mien, H. D. H. (2006). *Results of an experimental study about the effects of assessing Sixth Graders in Ho Chi Minh City who learned the reformed English language curriculum.* Paper presented at the Assessment for Improving Secondary and High School Students' Activeness, Ho Chi Minh City University of Pedagogy, Vietnam.

Milbrandt, M. K., Felts, J., Richards, B., & Abghari, N. (2004). Teaching-to-learn: A constructivist approach to shared responsibility. *Art Education, 57*(5), 19-33.

Militello, M., Bass, L., Jackson, K. T., & Wang, Y. (2013). How data are used and misused in schools: Perceptions from teachers and principals. *Education Sciences, 3*(2), 98-120.

Minh, N. T. T. (2007). Texbook evaluation: The case of English textbooks currently in use in Vietnam's upper-secondary schools: SEAMEO Regional Language Centre.

MOET. (2002). *34/2002/CT-BGDĐT: Direction for the central responsibilities of education in the School year 2002-2003*. Ha Noi.

MOET. (2003a). *38/2003/CT-BGDĐT: Direction for the central responsibilities of education in the School year 2003-2004*. Ha Noi: MOET.

MOET. (2003b). *50/2003/QD-BGD&DT: Decision on issuing the curriculum of English language and Informatics for primary school students.* Hanoi: MOET.

MOET. (2004). *25/2004/BGD&ĐT: Direction for the central responsibilities of education in the School year 2004-2005*. Ha Noi: MOET.

MOET. (2005a). *22/2005/CT-BGD&ĐT: Direction for the responsibilities of education in the School year 2005-2006*. Ha Noi: MOET.

MOET. (2005b). */BGD&ĐT-KT&KĐ: On objective tests of foreign languages in 2006*. Ha Noi: MOET.

MOET. (2006a). *16/2006/QĐ-BGDĐT: Decision on issuing the School education curriculum*. Ha Noi: MOET.

MOET. (2006b). *32/2006/CT-BGDĐT: Direction for the central responsibilities of pre-school, general, continuing, and professional education and pedagogic colleges and departments in School year 2006-2007*. Ha Noi: MOET.

MOET. (2006c). *40/2006/QĐ-BGDĐT: Decision on issuing regulations on assessing and ranking secondary school and high school students*. Ha Noi: MOET.

MOET. (2006d). *6912/BGDĐT-GDTrH: Guidelines on performing the responsibilities in secondary and high schools in the School year 2006-2007*. Ha Noi: MOET.

MOET. (2006e). *The school education curriculum - General issues (Attached with Decision 16/2006-QĐ-BGDĐT issued on May, 5th, 2006 by MOET)*. Ha Noi: MOET.

MOET. (2006f). *The school education curriculum for English language (Attached with Decision 16/2006-QĐ-BGDĐT issued on May, 5th, 2006 by MOET)*. Ha Noi: MOET.

MOET. (2007). *39/2007/CT-BGDĐT: Direction for the central responsibilities of pre-school, general, continuing, and professional education and pedagogic colleges and departments in School year 2007-2008*. Ha Noi: MOET.

MOET. (2008a). *47/2008/CT-BGDĐT: Direction for the central responsibilities of pre-school, general, continuing, and professional education in School year 2008-2009*. Ha Noi: MOET.

MOET. (2008b). *7475/BGDĐT-GDTrH: Guidelines on performing responsibilities in secondary and high schools in the School year 2008-2009*. Ha Noi: MOET.

MOET. (2009a). *264/BGDĐT-GDTrH: Organising conferences on reforming assessment to promote reform in teaching methods in Literature, History, and Civics in secondary and high schools*. Ha Noi: MOET.

MOET. (2009b). *287/TB-BGDĐT: Announcement about the conclusion of Deputy Minister Nguyen Minh Hien at the Conference on reforming assessment to promote teaching methods held by MOET in Can Tho on April 16-17, 2009*. Ha Noi: MOET.

MOET. (2009c). *4899/CT-BGDĐT: Direction for the central responsibilities of pre-school, general, continuing, and vocational education in the School year 2009-2010*. Ha Noi: MOET.

MOET. (2009d). *7394/BGDĐT: Guidelines on performing the responsibilities in secondary and high schools in the School year 2009-2010*. Ha Noi: MOET.

MOET. (2009e). *English curriculum for high schools*. Ha Noi: MOET.

MOET. (2010a). *4718/BGDĐT-GDTrH: Guidelines on performing the responsibilities in secondary and high schools in the School year 2010-2011*. Ha Noi: MOET.

MOET. (2010b). *8773/BGDĐT-GDTrH: Guidelines on designing tests*. Ha Noi: MOET.

MOET. (2010c). *English curriculum for primary schools*. Ha Noi: MOET.

MOET. (2010d). *Guidelines on designing tests (Attached with 8773/BGDĐT-GDTrH: Guidelines on designing tests issued by MOET on 30/12/2010)*. Ha Noi: MOET.

MOET. (2010e). *Professional development document on constructing tests and test banks for administrators and teachers: English language in high schools (Internal circulation document)*. Ha Noi: MOET.

MOET. (2011a). *30/2011/TT-BGDĐT: Circular on issuing the Continuing professional development programme for high school teachers*. Ha Noi: MOET.

MOET. (2011b). *58/2011/TT-BGDĐT: Circular on issuing the regulations on assessing and ranking secondary and high school students*. Ha Noi: MOET.

MOET. (2011c). *3398/CT-BGDĐT: Direction for the central responsibilities of kindergarten, general, continuing, and professional education in School year 2011-2012*. Ha Noi: MOET.

MOET. (2011d). *5358/BGDĐT-GDTrH: Guidelines on performing responsibilities in secondary and high schools in the School year 2011-2012*. Ha Noi: MOET.

MOET. (2011e). *The continuing professional development programme for high schools teachers (Attached to 30/2011/TT-BGDĐT: Circular on issuing the continuing professional*

development programme for high school teachers issued on 08/8/2011 by MOET). Ha Noi: MOET.

MOET. (2011f). *Regulations on assessing and ranking secondary and high school students (Attached to Circular 58/2011/TT-BGDĐT on 12/12/2011)*. Ha Noi: MOET.

MOET. (2012a). *2737/CT-BGDĐT: Direction for the central responsibilities of pre-school, general, continuing, and vocational education in the School year 2012-2013*. Ha Noi: MOET.

MOET. (2012b). *5289/BGDĐT-GDTrH: Guidelines on performing the responsibilities in secondary and high schools in the School year 2012-2013*. Ha Noi: MOET.

MOET. (2013). *5466/BGDĐT-GDTrH: Guidelines on performing responsibilities in secondary and high schools in the School year 2013-2014*. Ha Noi: MOET.

MOET. (2014). *4099/BGDĐT-GDTrH: Guidelines on performing responsibilities in secondary and high schools in the School year 2014-2015*. Ha Noi: MOET.

Mondada, L. (2012). Video recording as the reflecxive preservation and configuration of phenomenal features for analysis. In H. Knoblauch, B. Schnettler, J. Raab, & H.-G. Soeffner (Eds.), *Video analysis: Methodology and methods. Qualitative audiovisual analysis in Sociology* (pp. 51-68). Frankfurt am Main: Peter Lang.

Moni, K. B., van Kraayenoord, C. E., & Baker, C. D. (2002). Students' perceptions of literacy assessment. *Assessment in Education, 9*(3), 319-342.

Morgan, D. L. (1993). Qualitative content analysis: A guide to paths not taken. *Qualitative Health Research, 3*(1), 112-121.

Morgan, W., & Wyatt-Smith, C. (2000). Im/proper accountability: Towards a theory of critical literacy and assessment. *Assessment in Education: Principles, Policy & Practice, 7*(1), 123-142.

Morley, L. (2003). *Quality and power in higher education*. Maidenhead: SRHE and Open University Press.

Morse, J. M. (1994). Designing funded qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 220-235). Thousand Oaks, CA: SAGE.

Moye, M. J., Henkin, A. B., & Egley, R. J. (2005). Teacher-principal relationships: Exploring linkages between empowerment and interpersonal trust. *Journal of Educational Administration, 43*(3), 260-277.

Must, O., & Must, A. (2013). Changes in test-taking patterns over time. *Intelligence, 41*, 780-790.

Nagy, P. (2000). The three roles of assessment: Gatekeeping, accountability, and instructional diagnosis. *Canadian Journal of Education / Revue canadienne de l'éducation, 25*(4), 262-279.

National Assembly. (2000). *40/2000/QH10: Resolution of the National Assembly of the Socialist Republic of Vietnam on reforming the school education curriculum*. Ha Noi.

Neuman, W. L. (2014). *Social research methods: Qualitative and quantitative approaches*. Harlow: Pearson Education.

Newton, P., Tunison, S., & Viczko, M. (2010). The school principal's role in large-scale assessment. *Canadian Journal of Education Administration and Policy, 105*, 1-24.

Nga, D. H. (2006). *The status quo of assessing the learning outcomes of the sixth graders in Ho Chi Minh City who learned the reformed English language curriculum*. Paper presented at the Assessment for Improving Secondary and High School Students' Activeness, Ho Chi Minh City University of Pedagogy, Vietnam.

Ngan, M. Y., Lee, J. C.-K., & Brown, G. T. L. (2010). Hong Kong principals' perceptions on changes in evaluation and assessment policies: They're not for learning. *Asian Journal of Educational Research and Synergy, 2*(1), 36-46.

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), 199-218.

Nieto, S. (2010). *Language, culture, and teaching: Critical perspectives* (2nd ed.). New York: Routledge.

Nolen, S. B., Haladyna, T. M., & Haas, N. S. (1992). Uses and abuses of achievement test scores. *Educational Measurement: Issues and Practice, 11*(2), 9-15.

Nunan, D. (2003). The impact of English as a global language on educational policies and practices in the Asia-Pacific region. *TESOL Quarterly, 37*(4), 589-613.

Nunan, D. (2006). Task-based language teaching in the Asian context: Defining 'task'. *Asian EFL Journal, 8*(3). Retrieved from

O'Reilly, K. (2005). *Ethnographic methods*. New York: Routledge.

Ong, S. L. (2010). Assessment profile of Malaysia: High-stakes external examinations dominate. *Assessment in Education: Principles, Policy & Practice, 17*(1), 91-103.

Padgett, D. K. (1998). *Qualitative methods in social work research: Challenges and rewards*. Thousand Oaks, CA: SAGE.

Palmer, D. (2005). A motivational view of constructivist-informed teaching. *International Journal of Science Education, 27*(15), 1853-1881.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd. ed.). Thousand Oaks, CA: SAGE.

Pekkanli, I. (2010). Educational policies: Traditional foreign language classroom assessment in Turkish high schools. *Turkish Studies, 11*(3), 465-476.

Peterson, E. R., & Irving, S. E. (2008). Secondary school students' conceptions of assessment and feedback. *Learning and Instruction, 18*(3), 238-250.

Pham, L. A. (2013). A case study into English classroom assessment practices in three primary schools in Hanoi: Implications for developing a contextualized formative assessment practice framework. *Vietnam National University Journal of Foreign Language Studies, 29*(1), 1-16.

Poehner, M. E. (2011). Dynamic assessment: Fairness through the prism of mediation. *Assessment in Education: Principles, Policy & Practice, 18*(2), 99-112.

Pollard, A. (1990). Toward a sociology of learning in primary school. *British Journal of Sociology of Education, 11*(3), 241-256.

Prime Minister. (2001). *14/2001/CT-TTg: Direction for reforming the school education curriculum to execute the Resolution 40/2000/QH10 of the National Assembly*. Ha Noi.

Pryor, J., & Lubisi, C. (2002). Reconceptualising educational assessment in South Africa - Testing times for teachers. *International Journal of Educational Development, 22*, 673-686.

Prytula, M., Noonan, B., & Hellsten, L. (2013). Toward instructional leadership: Principals' perceptions of large-scale assessment in schools. *Canadian Journal of Education Administration and Policy, 140*, 1-30.

Qian, D. D. (2014). School-based English language assessment as a high-stakes examination component in Hong Kong: Insights of frontline assessors. *Assessment in Education: Principles, Policy & Practice, 21*(3), 251-170.

Remesal, A. (2007). Educational reform and primary and secondary teachers' conceptions of assessment: The Spanish instance, building upon Black and Wiliam (2005). *The Curriculum Journal, 18*(1), 27-38.

Renihan, P., & Noonan, B. (2012). Principals as assessment leaders in rural schools. *The Rural Educator, 33*(3), 1-8.

Richards, J. C., & Rogers, T. S. (2001). *Approaches and methods in language teaching*. New York, NY: Cambridge University Press.

Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher, 39*(5), 406-412.

Ross, S. J. (2008). Language testing in Asia: Evolution, innovation, and policy challenges. *Language Testing, 25*(1), 5-13.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119-144.

Sanger, J. (1996). *The complete observer? A field research guide to observation*. London: The Falmer Press.

Savignon, S. J. (2002). Communicative language teaching: Linguistic theory and classroom practice. In S. J. Savignon (Ed.), *Interpreting communicative language teaching: Contexts and concerns in teacher education* (pp. 1-27). New Haven: Yale University Press.

Scruggs, T. E., White, K. R., & Bennion, K. (1986). Teaching test-taking skills to elementary-grade students: A meta-analysis. *The Elementary School Journal, 87*(1), 68-82.

Segers, M., & Tillema, H. (2011). How do Dutch secondary teachers and students conceive the purpose of assessment? *Studies In Educational Evaluation, 37*(1), 49-54.

Seidman, I. (2006). *Interviewing as qualitative research: A guide for researchers in education and the social sciences* (3rd ed.). New York: Teachers College, Columbia University.

Serafini, F. (2001). Three paradigms of assessment: Measurement, procedure, and inquiry. *The Reading Teacher, 54*(4), 384-393.

Shen, J., Cooley, V. E., Reeves, P., Burt, W. L., Ryan, L., Rainey, J. M., & Yuan, W. (2010). Using data for decision-making: Perspectives from 16 principals in Michigan, USA. *International Review of Education, 56*(4), 435-456.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4-14.

Shepard, L. A. (2008). Formative assessment: Caveat emptor. In C. A. Dwyer (Ed.), *The future of assessment: Shaping, teaching, and learning* (pp. 279-303). Mahwah, NJ: Erlbaum.

Shing, L. W., & Fai, H. S. K. (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), 185-198.

Shute, V. J. (2008a). Focus on feedback. *Review of Educational Research, 78*(1), 153-189.

Shute, V. J. (2008b). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189.

Sireci, S. G., & Rios, J. A. (2013). Decisions that make a difference in detecting differential item functioning. *Educational Research and Evaluation, 19*(2-3), 170-187.

Smith, E., & Gorard, S. (2005). 'They don't give us marks': The role of formative feedback in student progress. *Assessment in Education, 12*(1), 21-38.

Sol, R. C. (2005). Subjectivity. In T. Honderich (Ed.), *The Oxford companion to philosophy* (2nd ed., pp. 900). New York, NY: OUP.

Solomon, P. (2002). *The assessment bridge: Positive ways to link tests to learning, standards, and curriculum improvement*. Thousand Oaks, California: Corwin Press, Inc.

Spicer, N. (2004). Combining qualitative and quantitative methods. In C. Seale (Ed.), *Researching society and culture* (pp. 293-303). Thousand Oaks, CA: SAGE.

Spradley, J. P. (1980). *Participant observation*. London: Wadsworth Thomson Learning.

Stake, R. E. (2006). *Multiple case study analysis*. New York: The Guilford Press.

Steenkamp, N., & Northcott, D. (2007). Content analysis in accounting research: The practical challenges. *Australian Accounting Review, 17*(3), 12-25.

Stiggins, R. (1999). Assessment, student confidence, and school success. *The Phi Delta Kappan, 81*(3), 191-198.

Stiggins, R. (2002). Assessment crisis: The absence of assessment FOR learning. *The Phi Delta Kappa, 83*(10), 758-765.

Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *The Phi Delta Kappa, 87*(4), 324-328.

Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. London: Routledge.

Suizzo, M. A. (2000). The social-emotional and cultural contexts of cognitive development: Neo-piagetian perspectives. *Child Development, 71*(4), 846-849.

Tan, K. (2011). Assessment for learning reform in Singapore - Quality, sustainable or threshold? In R. Berry & B. Adamson (Eds.), *Assessment reform in education: Policy and practice* (pp. 75-87). Dordrecht: Springer.

Tanner, H., & Jones, S. (2003). *Marking and assessment*. London: Continuum.

Taylor, S. J., & Bogdan, R. (1998). *Introduction to qualitative research methods: A guidebook and resource* (3rd ed.). New York, NY: John Wiley & Sons.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, CA.: SAGE.

Thinh, D. H. (2006). *The role of English in Vietnam's foreign language policy: A brief history*. Paper presented at the 19th Annual EA Education Conference, Australia.

Thuan, N. Q. (2005). Designing and developing the instrument to assess listening comprehension. *Science Journal of Vietnam National University - Hanoi, Foreign Languages, XXI*(1), [no page number].

Tierney, R. D. (2014). Fairness as a multifaceted quality in classroom assessment. *Studies In Educational Evaluation, 43*, 55-69.

Torff, B., & Sessions, D. (2009). Principals' perceptions of the causes of teacher ineffectiveness in different secondary subjects. *Teacher Education Quarterly, 36*(3), 127-148.

Triplett, C. F., & Barksdale, M. A. (2005). Third through sixth graders' perceptions of high-stakes testing. *Journal of Literacy Research, 37*(2), 237-260.

Tuan, N. P. (2006). *Assessing learning outcomes in the direction of improving high school students' activeness.* Paper presented at the Assessment for Improving Secondary and High School Students' Activeness, Ho Chi Minh City University of Pedagogy, Vietnam.

Tuyet, H. (2006). *Assessing learning outcomes in school education: Advances and inadequacies.* Paper presented at the Assessment for improving secondary and high school students' activeness, Ho Chi Minh City University of Pedagogy, Vietnam.

Valencia, S. W. (2011). Using assessment to improve teaching and learning. In S. J. Samuels & A. E. Farstrup (Eds.), *What research has to say about reading instruction* (4th ed., pp. 379-405). Newark, DE: International Reading Association.

Van, H. V., Hoa, H. T. X., Loc, D. N., Minh, D. T., & Tuan, N. Q. (2006). *English 10, Teacher's book*. Ha Noi: Education Publishing House.

Vandeyar, S., & Killen, R. (2007). Educators' conceptions and practice of classroom assessment in post-apartheid South Africa. *South African Journal of Education, 27*(1), 101-115.

Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 19-36). Mahwah, NJ: Lawrence Erlbaum Associates.

Watt, H. M. G. (2005). Attitudes to the use of alternative assessment methods in mathematics: A study with secondary mathematics teachers in Sydney, Australia. *Educational Studies in Mathematics, 58*(1), 21-44.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (2000). *Unobstrusive measures* (Rev. ed.). Thousand Oaks, CA: SAGE.

Weeden, P., Winter, J., & Broadfoot, P. (2002). *Assessment: What's in it for schools?* London: RoutledgeFalmer.

Wenger, E. (2009). A social theory of learning. In K. Illeris (Ed.), *Contemporaries theories of Learning: Learning theorists . . . in their own words* (pp. 209-218). New York: Routledge.

White, M. D., & Marsh, E. E. (2006). Content analysis: A flexible methodology. *Library Trends, 55*(1), 22-45.

Wolf, D., Bixby, J., Glenn, J., & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. *Review of Research in Education, 17*(31), 31-74.

Xi, X. (2010). How do we go about investigating test fairness? *Language Testing, 27*(2), 147-170.

Xia, Y. (2014). Language theories and language teaching - From traditional grammar to functionalism. *Journal of Language Teaching and Research, 5*(3), 559-565.

Yildirim, A. (2004). Student assessment in high-school social-studies courses in Turkey: Teachers' and students' perceptions. *International Review of Education, 50*(2), 157-175.

Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). London: SAGE.

Zane, T. W. (2009). Performance assessment design principles gleaned from constructivist learning theory (Part 1). *TechTrends, 53*(1), 81-88.

Zeidner, M. (1998). *Test anxiety: The state of the art*. New York: Kluwer Academic Publishers.

Zieky, M. (2013). Fairness review in assessment. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology* (Vol. 1. Test theory and testing and assessment in industrial and organizational psychology, pp. 359-376). Washington DC: American Psychological Association.

Appendices

Appendix A. Protocol for the Interviews with the Principals

PROTOCOL FOR THE INTERVIEWS WITH THE PRINCIPALS
(Interview questions for the principals)

Date:
Time:
Place:
Interviewee: (*pseudonym*)
Warm-up:
- Greet the interviewee.
- State the purpose of the interview: I would like to know about (a) your perceptions of assessment, (b) assessment in your school, (c) your expectations of the EFL teachers' assessment, and (d) your support of the EFL teachers in your school.
  Questions to be asked:
     1. What do you think about when you hear the word "assessment"? What does the term mean to you?
     2. Could you please tell me about assessment in your school?
     3. Why do you expect your EFL teachers to assess their students?
     4. How do you expect your EFL teachers to assess their students?
     5. What are the influences on the assessment practices used in your school?
     6. What do you think about your EFL teachers' knowledge and skills in assessment?
     7. What have you done to support them in assessment?
     8. How much freedom do your EFL teachers have in making decisions about assessment in their classrooms?

Appendix B. Protocol for the Interviews with the EFL Teachers

PROTOCOL THE INTERVIEWS WITH THE EFL TEACHERS
(Interview questions for the EFL teachers)

Date:
Time:
Place:
Interviewee: (*pseudonym*)
Warm-up:
- Greet the interviewee.
- State the purpose of the interview: I would like to know about (a) your perceptions of assessment, (b) your assessment practice, and (c) the influences on your implementation of assessment in your classrooms.
  Questions to be asked:
1. What do you think about when you hear the word "assessment"? What does the term means to you?
2. Please tell me about assessment in your school.
3. Please tell me about assessment in your classroom.
4. How do you think assessment should be carried out?
5. How do you implement assessment in your classroom?
6. Why do you assess your students?
7. What are the influences on assessment in your school?
8. What are the influences on assessment in your classroom?
9. What are the expectations around assessment of the Ministry of Education and Training?
10. What are the expectations around assessment of your principal?
11. What are the expectations around assessment of your students?
12. What are the expectations around assessment of the parents of your students?
13. How do you meet the expectations of these different groups?

Appendix C. Protocol for the Classroom Observations

PROTOCOL FOR THE CLASSROOM OBSERVATIONS

Date: Time: Place:
Class: Lesson: Teacher: (*pseudonym*)
Total number of students: Number of absentees:

Page 1: Map of the classroom (space, objects, teacher, students)
Page 2 (and more pages when necessary):

Time *(start and end time of assessment activities)*	**Descriptive notes** *(records of teacher's and students' words, activities, interactions, feelings, purposes of assessment)*		**Resources** *(handouts, task sheets, etc.)*
	Teacher	Students	
	(Plenty of space for notes)	*(Plenty of space for notes)*	

Last page: Reflections (the researcher's thoughts, hunches, and insights)

Appendix D. Protocol for the Interviews with the Students

PROTOCOL FOR THE INTERVIEWS WITH THE STUDENTS
(Interview questions for the students)

Date:
Time:
Place:
Interviewee: (*pseudonym*)
Warm-up:
- Greet the interviewee.
- State the purpose of the interview: I would like to know about (a) your perceptions of assessment, (b) assessment in your classroom, and (c) your expectation of the EFL teacher's assessment.

Questions to be asked:
1. What do you think about when you hear the word "assessment"? What does the term mean to you?
2. Could you please tell me about assessment in your classroom?
3. How does your EFL teacher assess your learning?
4. How do you want your EFL teacher to assess your learning?
5. What types of assessment would you prefer if you could choose the types of assessment?
6. Why do you prefer your EFL teacher to use these types of assessment?
7. Why does your EFL teacher assess your learning?

Appendix E. Consent Form for Interviewees

CONSENT FORM FOR INTERVIEWEES

Title: The contexts of assessment in EFL classrooms in two high schools in Vietnam

NOTE: This consent form will remain with the researcher for his records.

I agree to take part in the research project specified above. I have had the project explained to me, and I have read the *Participant Information Sheet*. I understand that agreeing to take part means that:

I agree to be interviewed by the researcher	☐	Yes	☐	No
I agree to allow the interview to be audio-taped	☐	Yes	☐	No
I agree to make myself available for a further interview if required	☐	Yes	☐	No

I understand that I will be given a transcript of data concerning me for my approval before it is included in the write up of the research.

I understand that my participation is voluntary, that I can choose not to participate in part or all of the project, and that I can withdraw at any stage of the project without being penalised or disadvantaged in any way.

I understand that any data that the researcher extracts from the interview for use in reports or published findings will not, under any circumstances, contain names or identifying characteristics.

Participant's name   Signature   Date

-----------------------   --------------------   ---------------

Appendix F. Consent Form for the Classroom Observations

CONSENT FORM FOR THE CLASSROOM OBSERVATIONS

Title: The contexts of assessment in EFL classrooms in two high schools in Vietnam

NOTE: This consent form will remain with the researcher for his records.

I agree to take part in the research project specified above. I have had the project explained to me, and I have read the *Participant Information Sheet*. I understand that agreeing to take part means that:

I agree to allow the researcher to observe my classrooms ☐ Yes ☐ No
I agree to allow my lessons to be audio-taped ☐ Yes ☐ No

I understand that my participation is voluntary, that I can choose not to participate in part or all of the project, and that I can withdraw at any stage of the project without being penalised or disadvantaged in any way.

I understand that any data that the researcher extracts from the observations for use in reports or published findings will not, under any circumstances, contain names or identifying characteristics.

Participant's name   Signature   Date

----------------------   --------------------   ---------------

Appendix G. Consent Form for Parents

CONSENT FORM FOR PARENTS

Title: The contexts of assessment in EFL classrooms in two high schools in Vietnam

NOTE: This consent form will remain with the researcher for his records.

I agree to allow my child to take part in the research project specified above. I have read the *Participant Information Sheet*. I understand that agreeing to allow my child to take part means that:

I agree that s/he will be interviewed by the researcher ☐ Yes ☐ No

I agree to allow the interview to be audio-taped ☐ Yes ☐ No

I agree to make my child available for a further interview if required ☐ Yes ☐ No

I understand that my child's participation is voluntary, that s/he can choose not to participate in part or all of the project, and that s/he can withdraw at any stage of the project without being penalised or disadvantaged in any way.

I understand that any data that the researcher extracts from the interview for use in reports or published findings will not, under any circumstances, contain names or identifying characteristics.

Student's name

Parent's name   Signature   Date

----------------------   --------------------   --------------

Appendix H. Participant Information Sheet

PARTICIPANT INFORMATION SHEET
(For principals, EFL teachers, and students)

Title: The contexts of assessment in EFL classrooms in two high schools in Vietnam

The purpose of the study

The study aims to examine high school principals' and EFL teachers' perceptions of the context of assessment in their school, how EFL teachers implement assessment in their classrooms, and how they negotiate the multiple purposes of assessment in their classrooms.

This study is being conducted by Le Nhan Thanh as part of the requirements for the PhD degree at the University of Queensland under the supervision of:

1. Associate Professor Dr. Christa van Kraayenoord
2. Associate Professor Dr. Karen Moni

Participation and withdrawal

Participation in this study is completely voluntary and you are free to withdraw from this study at any time without prejudice or penalty.

What is involved

Participants are asked to allow the researcher to interview and observe their classrooms.

Risks

There is no foreseeable risk.

Confidentiality and security of data

All data collected in this study will be stored confidentially. Only the researcher will have access to identified data. All data will be coded in a de-identified manner and subsequently analysed and reported in such a way that responses will not be able to be linked to any individual. The data you provide will only be used for the specific research purposes of this study.

Ethics Clearance and Contacts

This study has been cleared by one of the human ethics committees of the University of Queensland in accordance with the National Health and Medical Research Council's guidelines. If you would like to speak to an officer of the University not involved in the study, you may contact the Ethics Officer on 3365 3924.

If you would like to learn the outcome of the study, you can contact me at lenhanthanh2001@yahoo.com after 2014, and I will send you an Abstract of the study and findings.

Thank you for your participation in this study.

Le Nhan Thanh

Appendix I. Information Letter to the Head of the DOET

<div align="right">Date:</div>

INFORMATION LETTER TO THE HEAD OF THE DOET

To: Head of the Department of Education and Training of XYZ Province,

My name is Le Nhan Thanh. I am currently a PhD student at the School of Education, the University of Queensland, Australia. I am seeking permission to conduct the following PhD study.

The title of my dissertation is "The contexts of assessment in EFL classrooms in two high schools in Vietnam".

The study aims to examine principals', EFL teachers', and students' perceptions of the context of assessment in their school, how EFL teachers implement assessment in their classrooms, and how they negotiate the multiple purposes of assessment in their classrooms.

I can be contacted at (+84) 0912181716, or lenhanthanh2001@yahoo.com

My supervisors: 1. Associate Professor Dr. Christa van Kraayenoord
                       2. Associate Professor Dr. Karen Moni

This study has been cleared by one of the human ethics committees of the University of Queensland in accordance with the National Health and Medical Research Council's guidelines. If you would like to speak to an officer of the University not involved in the study, you may contact the Ethics Officer on 3365 3924.

To collect data for my study, I will interview the principals of two high schools, three EFL teachers and 36 students in each school. I will also observe classrooms of six EFL teachers in these high schools in a period of four months.

I am aware that my presence in the schools and classrooms may cause some inconvenience for the principals, teachers, and students, so I will try to minimise the possible inconvenience.

The principals, the EFL teachers and their students will participate in the study on a voluntary basis. They may withdraw at any time without any consequence.

The information about the schools, the principals, the teachers, and the students and the information that they provide will be kept confidential. Specifically, the name of the schools and the participants will not be used in any of my documents, including my dissertation.

In return for the principals' and teachers' participation in my study, I will provide the schools and teachers with some materials about assessment, and I will hold a workshop on assessment for the teachers in these schools if they request me to do so.

I do hope that you will give me permission to conduct my PhD in two high schools in your province.

Sincerely yours,

Le Nhan Thanh   Associate Professor Dr. Christa van Kraayenoord
PhD student   Principal supervisor

Appendix J. Information Letter to the Principals

INFORMATION LETTER TO THE PRINCIPALS

To: Principal of High School XYZ,

My name is Le Nhan Thanh. I am currently a PhD student at the School of Education, the University of Queensland, Australia. I am seeking permission to conduct the following PhD study.

The title of my dissertation is "The contexts of assessment in EFL classrooms in two high schools in Vietnam".

The study aims to examine principals', EFL teachers', and students' perceptions of the context of assessment in their school, how EFL teachers implement assessment in their classrooms, and how they negotiate the multiple purposes of assessment in their classrooms.

I can be contacted at (+84) 0912181716, or lenhanthanh2001@yahoo.com

My supervisors: 1. Associate Professor Dr. Christa van Kraayenoord

                    2. Associate Professor Dr. Karen Moni

This study has been cleared by one of the human ethics committees of the University of Queensland in accordance with the National Health and Medical Research Council's guidelines. If you would like to speak to an officer of the University not involved in the study, you may contact the Ethics Officer on 3365 3924.

To collect data for my study, I will conduct one interview with you about assessment in your school. I will also observe 21 classrooms of three EFL teachers in your school and interview them in a period of four months. In addition, I will interview six groups of 36 students in the EFL classes about their perceptions of assessment in their classroom.

I am aware that my presence in your school and classrooms may cause some inconvenience for the teachers and students, so I will try to minimise the possible inconvenience.

You and the EFL teachers and students will participate in the study on a voluntary basis. You and the EFL teachers and students may withdraw at any time without any consequence.

The information about yourself, your school, teachers, and students and the information that you and the teachers and students provide will be kept confidential. Specifically, your name, the school's, teachers', and students' name will not be used in any of my documents, including my dissertation.

In return for the teachers' and students' participation in my study, I will provide your school and teachers with some materials about assessment, and I will hold a workshop on assessment for your teachers if you request me to do so.

I do hope that you will give your permission to collect data for my study in your school.

Sincerely yours,

Le Nhan Thanh   Associate Professor Dr. Christa van Kraayenoord

PhD student   Principal supervisor

Appendix K. Information Letter to the Principals as Participants

<div align="right">Date:</div>

INFORMATION LETTER TO THE PRINCIPALS AS PARTICIPANTS

To: Mr/Ms XYZ,

My name is Le Nhan Thanh. I am currently a PhD student at the School of Education, the University of Queensland, Australia.

The title of my dissertation is "The contexts of assessment in EFL classrooms in two high schools in Vietnam".

The study aims to examine principals', EFL teachers', and students' perceptions of the context of assessment in their school, how EFL teachers implement assessment in their classrooms, and how they negotiate the multiple purposes of assessment in their classrooms.

I can be contacted at (+84) 0912181716, or lenhanthanh2001@yahoo.com

My supervisors: 1. Associate Professor Dr. Christa van Kraayenoord
                       2. Associate Professor Dr. Karen Moni

This study has been cleared by one of the human ethics committees of the University of Queensland in accordance with the National Health and Medical Research Council's guidelines. If you would like to speak to an officer of the University not involved in the study, you may contact the Ethics Officer on 3365 3924.

To collect data for my study, I will conduct one interview with you about (a) your perceptions of assessment, (b) assessment in your school, (c) your expectation of the EFL teachers' assessment, and (d) your support to the EFL teachers in your school.

The information about yourself and your school and the information that you provide will be kept confidential. Specifically, your name and your school's name will not be used in any of my documents, including my dissertation.

I am grateful for your agreement to participate in my study.

Sincerely yours,

Le Nhan Thanh   Associate Professor Dr. Christa van Kraayenoord
PhD student   Principal supervisor

Appendix L. Information Letter to the EFL Teachers

INFORMATION LETTER TO THE EFL TEACHERS

To: Mr/Ms XYZ,

My name is Le Nhan Thanh. I am currently a PhD student at the School of Education, the University of Queensland, Australia. I am seeking your agreement to participate in my PhD study.

The title of my dissertation is "The contexts of assessment in EFL classrooms in two high schools in Vietnam".

The study aims to examine principals', EFL teachers', and students' perceptions of the context of assessment in their school, how EFL teachers implement assessment in their classrooms, and how they negotiate the multiple purposes of assessment in their classrooms.

I can be contacted at (+84) 0912181716, or lenhanthanh2001@yahoo.com

My supervisors: 1. Associate Professor Dr. Christa van Kraayenoord
                    2. Associate Professor Dr. Karen Moni

This study has been cleared by one of the human ethics committees of the University of Queensland in accordance with the National Health and Medical Research Council's guidelines. If you would like to speak to an officer of the University not involved in the study, you may contact the Ethics Officer on 3365 3924.

To collect data for my study, I will observe seven of your classrooms in a period of four months. I will also conduct one interview with you and two focus group interviews with two groups of your students, each with six students.

I am aware that my presence in your classrooms may cause some inconvenience for you and your students, so I will try to minimise the possible inconvenience.

You will participate in the study on a voluntary basis. You may withdraw at any time without any consequence.

The information about yourself, your school, and your students and the information that you and your students provide will be kept confidential. Specifically, your name, your school's and students' name will not be used in any of my documents, including my dissertation.

In return for your participation in my study, I will provide you with some materials about assessment and pay you for spending time on the interviews (personally negotiated).

I am grateful for your agreement to participate in my study.

Sincerely yours,

Le Nhan Thanh   Associate Professor Dr. Christa van Kraayenoord
PhD student   Principal supervisor

Appendix M. Permission to Conduct Research from Head of the DOET

Date:

PERMISSION TO CONDUCT RESEARCH FROM THE HEAD OF THE DOET

To Whom It May Concern.
The purpose of this letter is to inform you that I give Mr. Le Nhan Thanh permission to conduct the research titled **The contexts of assessment in EFL classrooms in two high schools in Vietnam** in [Name of School(s)]

Sincerely,

[Signature]

[Name of Signatory]
[Title of Signatory]

Appendix N. Table Used in Data Analysis

Reference	Text Segment	Category	Theme	Cluster
MOET. (2010e). *Professional development document on constructing tests and test banks for administrators and teachers: English language in high schools (Internal circulation document).* Ha Noi: MOET.	The reform in assessment was "a requirement that needs to be fulfilled when the reform in the methods of teaching and learning and the reform in education [in general] are undertaken" (p. 4).	Integral part of reform in school education	The need for reform in assessment	Reform in assessment
MOET. (2010e). *Professional development document on constructing tests and test banks for administrators and teachers: English language in high schools (Internal circulation document).* Ha Noi: MOET.	". . . reform in methods of teaching and learning and reform in assessment are the key solutions to increasing the quality of teaching and learning in particular and the quality of comprehensive education in general" (pp. 11-12).	Promoting changes in methods of teaching and learning	Policy-makers' expectations of positive changes from reform in assessment	Reform in assessment
MOET. (2010e). *Professional development document on constructing tests and test banks for administrators and teachers: English language in high schools (Internal circulation document).* Ha Noi: MOET.	"When reform in assessment is implemented, it is necessary to ensure the requirement of objectivity, accuracy, [and] fairness" (p. 6)	Ensuring accuracy, objectivity, and fairness in assessment	Policy-makers' expectations of positive changes from reform in assessment	Reform in assessment
Interview with Principal A	"Marks must be the accurate measures of the students' learning outcomes."	Tests as a technology for measuring knowledge	Concerns about technical issues	Concerns about technical and ethical issues
Interview with Principal A	". . . the expectations in assessment in this school are quite high. . . . the teachers in this school require more of their students"	Ethical issues associated with the teachers' test construction	Concerns about ethical issues	Concerns about technical and ethical issues