Table of contents

List of Tables

List of Figures

Chapter 1. Introduction

This thesis reports on multiple case studies of three EFL teachers' classroom assessment practices in their tertiary-level oral English courses in China. In this chapter, after a clarification of the terms used in this study, the author describes the research background and the purpose of the present study, followed by the significance of this study. This chapter ends with a brief overview of this thesis.

1.1 Clarifying the terms

Generally speaking, *classroom assessment* (hereafter CA) refers to assessment conducted in classroom settings. The word *classroom* emphasizes that it refers to those assessment practices conducted in classroom settings in contrast to external standardized testing (Leung, 2005). By using the word *assessment*, this term refers to the process of collecting evidence and making judgment about students' achievement rather than about programs, in which case *evaluation* is often used (Harlen, 2007). Since this thesis is mainly concerned with how teachers assessed their students in classroom contexts, the term *classroom assessment* instead of *classroom evaluation* is adopted.

Moreover, the term *assessment* instead of *testing* is used in this thesis because *assessment* has broader connotations. The term *assessment* is often used "as a general umbrella term to cover all methods of testing and assessment," and *testing* often refers to the construction and administration of formal or standardized tests (Clapham, 2000, p. 150). To avoid confusion, in this thesis, the term *assessment* was used to cover all the possible assessment methods, from formal tests to informal alternative assessment methods, and the term *test* was used to refer to those formally constructed and administered assessment methods such as a final test.

Presently, while some researchers use CA interchangeably with *formative assessment* (hereafter FA) (Stiggins & Conklin, 1992), some other researchers use it to include not only formative but also summative functions of CA (Cheng, Rogers, & Hu, 2004; Cheng, Rogers, & Wang, 2008; Cheng & Wang, 2007; Rea-Dickins, 2001). *Summative assessment* (hereafter SA) is usually distinguished from FA along two

dimensions. Along the time dimension, FA is usually administered in the course of a unit of instruction while SA is administered at the end of a unit of instruction; along the purpose dimension, FA is usually for improving teaching and learning while SA is usually for categorizing students' performances or certification (Cizek, 2010). As Brookhart (2004) has pointed out, "What British authors call formative assessment or teacher assessment is often called classroom assessment in the United States literature, although classroom assessment in the United States does include some aspects of assessment that are more properly considered summative—for example, grading" (p.429). To overcome such confusion, in this thesis, the term CA with its broader connotation is adopted, that is, this term encompasses both the formative and the summative functions of teachers' CA practices. More detailed discussion can be found in section 2.2.

1.2 Research background

While assessment has always been an integral part of courses or instruction, extensive research on teachers' CA practices can be traced back to only about three decades ago when, as a way to respond to the concern about "the wholesale reliance on mandated multiple-choice tests" (Calfee & Masuda, 1997, p. 71), a large volume of published studies appeared that investigated the impact of assessment practices in schools and classrooms on students (c.f., Crooks, 1988; Natriello, 1987). This effort to strike a balance between teacher-based CA and external standardized tests continued in the 1990s. In 1998, Black and Wiliam's seminal paper (1998a) brought about a shift in focus in later research. This substantial literature review not only pointed out the features of teachers' CA practices that can enhance student learning or improve teacher instruction that may consequently improve student learning, but also reported a big failure in implementing such beneficial assessment practices in actual classroom contexts. Therefore, since then there has been a large and growing body of empirical studies as well as pedagogical texts on the nature and implementing strategies of such beneficial assessment practices, and on the impact of such assessment practices on students (Andrade & Cizek, 2010; Brookhart & DeVoge, 1999; Brookhart & DeVoge, 2000; Brookhart & Durkin, 2003; Gardner, 2006b; McMillan, 2007a, 2007c; Popham, 2008a, 2008b, 2011). These trends emphasize teacher-made assessment conducted in

the classroom settings and are often linked directly to teaching with an intention to enhance teaching and learning of all students rather than just for ranking and selecting purposes.

Research on second/foreign language teachers' CA practices has showed a similar developmental pattern over the past few decades. In the 1980s, the language assessment field was dominated by language testing research where the focus was on how to look at language ability, and how to design, develop, and validate language tests, especially large-scale standardized tests that could assess the specific characteristics of a given population of test takers (Bachman, 1990, 1991; Douglas & Chapelle, 1993). This dominance greatly influenced language teachers' CA practices, which generally followed the standardized testing format that relied largely on discrete-point items (Rhodes, Rosenbusch, & Thompson, 1997). Alderson's (1986) paper directed some researchers' attention to the washback effects of language tests on classroom teaching and learning and prompted a number of washback studies subsequently (Alderson & Wall, 1993; Cheng, 2005; Cheng, Watanabe, & Curtis, 2004; Green, 2007; Wall, 2005). Although the washback studies paid a lot of attention to the perceptions and behaviours of both teachers and students in classroom contexts, their starting point was usually a test, very often an externally carefully designed high-stake test rather than teachers' actual CA practices. The idea behind this group of studies was that tests direct teaching and learning.

Around the same period of time, as responses to the dissatisfaction with standardized testing, some researchers began to explore alternative assessment methods, such as the ground-breaking work on portfolio assessment by Liz Hamp-Lyons and William Condon (Cordon & Hamp-Lyons, 1991; Hamp-Lyons & Condon, 1993), or outcome-based assessment (Brindley, 1998, 2001). However, alternative assessment research was mainly concerned with the development of assessment tools rather than the interfaces between language assessment and classroom curricula and pedagogy (Rea-Dickins, 2008, p. 257), and outcome-based assessment research mainly focused on developing "a range of criterion-referenced procedures and instruments which can be used to monitor and assess achievement" (Brindley, 1998, p. 47). This shift in focus to regard assessment as a servant to rather than a master of classroom teaching and

learning in the language assessment field was further clarified and emphasized by McNamara (2001) when he pointed out that the validity demands of language testing researchers could not serve the needs of classroom-based assessment and advocated to expand the notion of language assessment to encompass "any deliberate, sustained and explicit reflections by teachers (and by learners) on the qualities of a learner's work" (p.343) so as to make language assessment research more accountable to the needs of teachers and students (pp. 343-345). This idea coincided with the idea advocated by Black and Wiliam (1998a) in that both called for more research on CA practices that serve rather than control teaching and learning. Since then, there has been an increasing number of studies and discussion papers on the nature of CA in language learning classrooms, on language teachers' actual CA practices, and on the impacts of existing or newly-designed CA practices on teaching and learning (e.g., Carless, 2011; Carless, Joughin, Liu, & Associates, 2006; Colby-Kelly & Turner, 2007; Rea-Dickins, 2001, 2004, 2006).

While CA research has received more and more emphasis over the years, three research gaps have emerged. First, while CA is a complex system involving many components and many variables (Harlen, 2007), few empirical studies have been conducted to reveal the complexity of this system as a whole. Second, while a lot of research has been conducted to reveal the features of FA, little attention is given to SA and the relationship between FA and SA, although some researchers advocated that all assessment, both formative and summative, should be conceptualized as assessment for learning and synergy should be achieved between FA and SA (Carless, 2008; Davison & Leung, 2009; Harlen, 2005; Taras, 2005). Third, "the unfolding research agenda for classroom language assessment has been shaped in EAL/ESL rather than in foreign language learning contexts" (Rea-Dickins, 2008, p.258). Considering that China is a unique country, famous for its large number of English language learners (Jin & Cortazzi, 2006) and its "exam-oriented" learning culture (Cheng, 2008; Han & Yang, 2001), and the EFL contexts in China are still very much under-researched regarding CA research, an investigation on EFL teachers' CA practices in a Chinese EFL context should enrich the current knowledge of CA theories and practices.

1.3 Study purpose

This study investigated the CA practices of three tertiary-level oral English teachers in China. It was intended to gain an in-depth, contextualized, and comprehensive understanding of the teachers' CA practices in their oral English classrooms, so as to identify the components and variables involved in their CA practices and how these elements functioned as a whole in naturalistic settings over time. Following McNamara's (2001) suggestion that language assessment research should be made "answerable to the needs of teachers and learners" (p.346), this study not only examined the actual implementation of the teachers' CA practices, but also probed the teachers' explanations of their assessment practices, and their students' perceived impacts.

In addition, as the current understanding of CA has been extended to include not only formal assessment practices such as tests and quizzes but also classroom interaction episodes that involve assessment (e.g., Leung & Mohan, 2004; Rea-Dickins, 2006), which might remain oblivious to teachers who have not received systematic training on CA, the present study also attempted to highlight both the teachers' recognized CA practices and those unrecognized by the participants, so as to present a comprehensive picture of the teachers' CA practices.

1.4 Significance of the study

This study has both theoretical and practical significance. Theoretically, the study findings generated from this unique educational context will broaden and strengthen the understanding of the nature of CA. China is a unique country in terms of English language learning and assessment. With the acceleration of globalization and rapid development of science and technology in China, English is taking an increasingly important position in both the educational system and the whole society. In the field of education, English is an obligatory subject from primary school till post-graduate study. For all the significant examinations in one's formal education, such as high school entrance examination, college entrance examination and post-graduation entrance examination, English is a core examination subject. "Apart from English as an academic requirement, English skills are tested for all those seeking promotion in

governmental, educational, scientific research, medical, financial, business and other government-supported institutions" (He, 2001, cited in Cheng & Curtis, 2010a, p. 8). Accordingly, the English language assessment practice is dominantly summative (Carless, 2011; Cheng & Curtis, 2010b). Therefore, situating this study in a Chinese EFL context may reveal some unique principles and practices that may be in sharp contrast with those found in "Anglophone settings" (UK, US, Australia, New Zealand) (Carless, 2011, p.3) where current theories and good practices about CA have originated and prospered. Practically, since assessing students in the classroom is an important skill for teachers (Inbar-Lourie, 2008; Leung, 2004), the findings of this study will reveal the participant teachers' assessment literacy, which can serve as the basis for future intervention studies aiming at enhancing teachers' assessment literacy.

This study also has personal significance. Since I became an EFL teacher in 1997 after I obtained my first master's degree in Applied Linguistics, I have developed an interest in language testing. Later, my second master's degree in Education turned my attention from language testing to CA in the classroom. Over the years, while I have become increasingly aware of the importance and power of CA in classroom teaching and learning, I have also become increasingly clear of the shortcomings of my previous studies. Therefore, this PhD study has been an extension of my research interest and an opportunity for me to accomplish what I have always wanted to do.

1.5 Overview of the thesis

This thesis consists of eight chapters. Chapter 1 lays out the background and study purpose for this study. Chapter 2 presents a literature review of the nature of CA and the existing relevant models on CA, followed by a critique of the empirical studies conducted in L2 assessment field. A framework for the present study is then proposed. Chapter 3 presents the methodology, which includes the overall research design and the specific research strategies used in the study. Chapters 4 to 6 present the findings from the three cases, and such findings are compared / contrasted and then discussed against the existing literature in chapter 7. Chapter 8 concludes the study and points out its implications.

Chapter 2. Literature Review

In four parts, this chapter begins with an overview of the historical roots of CA. Part two explores the nature and scope of CA, discussing its multiple dimensions and multiple functions in educational contexts. Part three discusses the existing models/ frameworks of CA, and part four is a critique of the empirical studies on CA in the second language (L2) assessment field. This chapter ends with a framework for the present study.

2.1 The origins of CA

CA, sitting at the intersection between assessment, classroom teaching, and student learning (Cizek, 1997), can be traced back to *formative evaluation* (Scriven, 1967) and *criterion-referenced measurement* (Glaser, 1963), both of which were originally related to curriculum evaluation rather than assessment of student achievement.

One source of CA is found in Scriven's (1967) concept of *formative evaluation*. Working in the context of curriculum evaluation, Scriven (1967) pointed out that evaluation "can and usually should play several roles" (p.41). In addition to evaluating "the final product" of an educational program (summative role), it should also serve as "an important part of the process of curriculum development" (formative role) (p.41). In his opinion, summative evaluation occurred at the end of a program to determine its effectiveness or worth, while formative evaluation occurred while a program was still being implemented to make improvement decisions regarding the program. Later, Bloom and his colleagues borrowed the concept of *formative evaluation* and used it to measure student learning (Bloom, Hastings, & Madaus, 1971). Today, the essence of this formative-summative distinction has been retained and has been applied to assessment in schools. In other words, Scriven's concept of *formative evaluation* contributed to today's understanding of the formative functions of CA, or FA.

Another source of CA can be traced back to Glaser's (1963) concept of *criterion-referenced measurement*. Out of concern for an appropriate way to evaluate instructional programs, Glaser proposed a distinction between two approaches to interpreting students' test performances: *norm-referenced measurement* (hereafter

NRM) and *criterion-referenced measurement* (hereafter CRM). NRM aims to determine an examinee's relative status in relation to all the other examinees who take the same test; and CRM aims to provide a clear picture of what a student has actually acquired against predetermined descriptions of knowledge and/or skills to be learned.

The proposal of CRM was significant for present-day CA practices because CRM emphasizes the link between assessment and curriculum. Just as CRM stresses clear behavioural objectives, an effective FA strategy pointed out by Wiliam and Thompson (2008, p. 64) is that teachers should clarify and ensure students' understanding of learning intentions and criteria for success. Since CRM is concerned with interpreting student achievement, it should be regarded as one basis of today's FA practices in classroom settings, because FA goes one step further to use such interpretations to modify teaching and improve learning. As suggested by Popham (2011), the bulk of a teacher's CA practices "will be better served by criterion-referenced rather than norm-referenced interpretations because criterion-referenced interpretations tend to provide teachers with a clearer picture of what it is students can or can't do" (p.57).

Although both concepts were proposed in the 1960s, it was not until the 1980s that they began to gain increasing attention and importance, which led to an increase in CA research in the 1990s.

In the 1970s and 1980s in the US, in response to an increasing demand for educational accountability, CRM was adopted in many state-wide achievement tests measuring whether students had acquired the basic skills before they could receive their high school diplomas or be advanced to higher grade levels (Popham, 1994, 2011). However, these CRM objectives were measured by means of the same kind of discrete-point multiple-choice items that had been used for a long time in standardized NRM tests. Since the mid-1980s, with a growing dissatisfaction from the general public and the government officials with the traditional standardized tests (Shepard, 1989), there was a growing interest in forms of assessment other than standardized tests (e.g., Herman, Aschbacher, & Winters, 1992; Mitchell, 1992; Wiggins, 1989) and emphasis on teachers as assessors instead of agents for external professional testing agencies (cf., Crooks, 1988; Natriello, 1987). Similar trends also appeared in the

language assessment field (Hamayan, 1995; Huerta-Macias, 1995; Law & Eckes, 1995; O'Malley & Pierce, 1996). As Calfee and Masuda (1997) concluded in their review:

> [T]he capacity of classroom teachers, as members of a professional community, to assess student achievement by means of less than fully standardized methods and to connect these assessments to the ongoing improvement of the instructional program is critical for meeting the national agenda of ensuring high-quality education for all students. (p.70)

In the 1990s, with the increasing attention given to teachers' CA practices, especially in relation to teaching and learning, there came a paradigm shift from a testing culture to an assessment culture (e.g., Gipps, 1994). This was further boosted by Black and Wiliam's seminal review paper (1998a). Through a review of about 250 empirical studies relating to CA, one significant finding was that "attention to formative assessment can lead to significant learning gains" (p.17). This powerful finding and its dissemination (Black & Wiliam, 1998b, 1998c) brought about a shift in focus regarding CA research, from the idea that assessment was useful only for appraising student learning to the idea that assessment could inform instruction and learning. In other words, the new focus put more emphasis on how assessment could be used to improve teaching and enhance learning than on how to measure student learning according to the traditional psychometric procedures. This shift also occurred in the language assessment field (McNamara, 2001). This new focus led to a proliferation of new terms, which will be discussed in the next section.

2.2 Defining CA

CA is a hot topic in present-day educational research (cf., Andrade & Cizek, 2010; Brookhart, 2004; Gardner, 2006b; Harlen, 2007; Popham, 2011), and is also gaining increasing attention in the L2 assessment field (cf., Davison & Cummins, 2007a; Davison & Leung, 2009; Rea-Dickins, 2007a). The dynamics of this research area is marked by a number of similar terms that are being used these days, such as *classroom assessment* (CA), *formative assessment* (FA), *assessment for learning* (AfL), *learning-oriented assessment* (LOA), *teacher-based assessment* (TBA), *dynamic assessment* (DA), and *alternative assessment* (AA). In this section, major terms used to talk about teachers' assessment practices in their classroom contexts (from both the

education field in general and language assessment field in particular) and their respective emphases will be presented and discussed, and then CA will be defined for the present study.

2.2.1 Related terms of CA

Classroom assessment environments and events

CA is not a new term and its meaning has been expanded over the years, though some inconsistencies still exist. It started to appear in the educational research field in the early 1990s in the US (Angelo & Cross, 1993; Stiggins & Conklin, 1992), when it was used to refer to "those assessments developed and used by teachers in the classroom on a day-to-day basis" (Stiggins & Conklin, 1992, p.1). It was used as a contrast to external standardized tests, and the contrasts lay in three aspects: timing (on a day-to-day basis rather than at the end of an instructional program), context (used as an everyday classroom activity rather than administered to a large number of testees in a standardized manner), and agent (developed and used by teachers rather than by external agencies and professional testers).

Stiggins and Conklin (1992) also suggested the concept of *classroom assessment environment*, because they found, through an observational study of three 6th-grade classrooms, that teachers' assessment purposes and practices varied along eight dimension[1], which, taken together, seemed to describe an assessment environment. Brookhart in her literature review (1997) further developed this concept and also proposed a new concept: *classroom assessment event*. She argued that, in classroom settings, teachers conduct many classroom assessments and communicate their expectations through the way they conduct these assessments. Students, on the other hand, experience specific expectations each time a task is assigned and learn how well they meet those expectations through teachers' feedback. These tasks are referred to as *classroom assessment events*. Because there are many of these classroom assessment events in typical classrooms, the overall sense of expectation that is communicated by

[1]Dimensions of the classroom assessment environment: 1) assessment purpose, 2) methods used to assess achievement, 3) criteria used in selecting assessment methods, 4) quality of assessment, 5) feedback, 6) teacher as assessor, 7) teacher's perceptions of the students, 8) assessment-policy environment (Stiggins& Conklin, 1992, p.80).

all of the assessment events is referred to as the *classroom assessment environment*. Brookhart (2003) also pointed out three essential features of CA, that is, it is context-dependent, it has an inextricable relationship with instruction, and it can perform simultaneous formative and summative functions.

It should be noted that *CA environment* and *CA event* emphasized either the factors that might shape a teacher's CA practices (Brookhart, 2003; Stiggins & Conklin, 1992) or students' perceived impacts of teachers' CA practices (Brookhart, 1997), while little attention was given to what assessment actually is. In other words, this concept focuses on "classroom" and took "assessment" for granted. In these definitions, CA more or less is equated to assessment methods.

In contrast, Cizek (1997) focused on "assessment" rather than "classroom". His definition specified not only the processes involved in assessment but also the functions that assessment should play in classroom contexts. CA is "the planned process of gathering and synthesizing information relevant to the purposes of (a) discovering and documenting students' strengths and weaknesses, (b) planning and enhancing instruction, or (c) evaluating progress and making decisions about students" (Cizek, 1997, p. 10). This definition pointed out that CA is a planned process involving three steps: gathering information about students, synthesizing the information collected, and making use of the information collected; and it can be used both formatively and summatively "with the primary beneficiary being the student" (ibid., p.10). It should be noted that in Cizek's definition, the word "purpose" means more or less the same as "function" or "use", all of which can be used to refer to how the judgment has been used, the third step of the assessment process. In this thesis, the three words are also used interchangeably to refer to the third step of the assessment process.

In her definition of CA, Shepard (2000) also made a contrast with standardized objective tests. CA is "not the kind of assessment used to give grades or to satisfy the accountability demands of an external authority, but rather the kind of assessment that can be used as a part of instruction to support and enhance learning" (p.4). Her definition is narrower than Cizek's (1997) in that it focuses on formative functions of

11

CA only, that is, to make instruction more effective so as to promote learning. In this sense, it is similar to such terms as FA or AfL (Black, Harrison, Lee, Marshall, & Wiliam, 2003; Cowie & Bell, 1999; Rea-Dickins & Gardner, 2000; Torrance, 1993; Torrance & Pryor, 1998), and instructional assessment or instructionally-oriented assessment (Airasian & Russell, 2008; Popham, 2008a) in that they all emphasize that CA is an integral part of classroom teaching and learning and its major function is to provide formative feedback for teaching and learning so as to enhance the quality of teaching and promote student achievement.

The definitions of this term from many textbooks on CA (e.g., Airasian & Russell, 2008; Banks, 2005; Brookhart & Nitko, 2008; Butler & McMunn, 2006; McMillan, 2007a; Popham, 2011; Stiggins, 2008) generally consider assessment in its general sense as a process involving collecting, interpreting, and using information collected, and naturally CA refers to such kind of processes in the classroom contexts. Moreover, the comprehensive information provided in these books about how to prepare and use various CA methods, how to assign grades and interpret the results, how to assure the quality of CA, and how to use CA to evaluate their own instructional effectiveness, all indicate that CA can take various forms and serve multiple purposes. Furthermore, CA is a planned and fairly formal process, unlike the interactive FA proposed by Cowie and Bell (1999), which takes the form of spontaneous interactions between teachers and students during the ongoing classroom exchanges with the aim of extending student learning or resolving students' difficulties.

Overall, in the general education field, CA is a term mainly used in US contexts. Its meaning has expanded from referring to possible assessment methods used in classroom contexts to a planned process involving three steps: collection, evaluation, and use of information about student learning, which may take various forms and serve multiple purposes with the ultimate goal to enhance student learning. For some researchers, CA is the same as FA, but for others, CA incorporates both FA and SA, as Brookhart (2004) commented.

Teacher-based assessment (TBA)

Although *Teacher-based assessment* (TBA) was first used in Davison's 2004 article, its full definition only came in Davison and Leung's 2009 paper (Davison & Leung, 2009). In their definition, they contrasted TBA with external standardized testing to emphasize the contextual features of TBA. TBA means

> [A] more teacher-mediated, context-based, classroom-embedded assessment practice, explicitly or implicitly defined in opposition to traditional externally set and assessed large-scale formal examinations used primarily for selection and/or accountability purposes. (ibid, p. 395)

Moreover, Davison and Leung also pointed out that TBA is a synonym of "*alternative assessment*, *classroom* and/or *school-based assessment*, *formative assessment*, and more recently, *assessment for learning*" (italicized original) (ibid, p.395), and "the goal of TBA is to improve student learning" (ibid, p.401).

In general, TBA stresses two points: the classroom contexts of assessment where teachers play an important role, and the learning-oriented function of assessment. It upholds the belief that all assessment is assessment for learning and includes "not only the formal planned moments when students undertake an assessment task but also the far more informal, even spontaneous moments when teachers are monitoring student group work and notice one student speaking more confidently or another failing to take an offered turn" (ibid, p. 401).

Formative assessment (FA)

The term FA also has a long history and its meaning has evolved over the years. As previously noted, the earliest effort to distinguish between formative and summative roles of evaluative practices is conventionally attributed by many researchers to Michael Scriven in the field of curriculum evaluation (e.g., Cizek, 2010; Taras, 2005). However, Carless (2011) argued that more credit should be given to Cronbach's (1964) paper in which Cronbach suggested that evaluation for improvement was more useful than evaluation for appraisal and Scriven's (1967) paper was only "a critique of, and response to, Cronbach" (Carless, 2011, p. 28). To counterbalance Cronbach's

preference for formative evaluation, Scriven suggested a separation of the two roles of evaluation: for curriculum development and for the "evaluation of the final product of this educational process" (Scriven, 1967, p. 41), both of which could be based on the same information. This idea foreshadows recent discussion on the relationship between FA and SA (e.g., Brookhart, 2010; Rea-Dickins & Gardner, 2000; Taras, 2005). Citing Scriven's 1991 paper, Carless also pointed out that Scriven "cautions against formative evaluation being done informally, and notes that it needs the same rigor as a good summative evaluation" (Carless, 2011, p. 29). It can be seen that originally formative program evaluation was defined from the perspective of the purpose of evaluation and that it was seen as needing to be formally planned.

While Scriven's concept of formative evaluation was applied only to program evaluation, it was Bloom and his colleagues (Bloom et al., 1971) who first applied this concept to student learning, which laid the foundation for today's FA research. They defined formative evaluation as "the use of systematic evaluation in the process of curriculum construction, teaching and learning for the purpose of improving any of these three processes" (p.117). Their formative evaluation took the form of progress tests, used at the end of a short unit, to determine student mastery of that unit, and then followed by corrective instruction designed to help students remedy their learning errors that were identified. Although this traditional conception of FA is rather "restricted" (Carless, 2011, p.90), it clearly indicates the uses of formative evaluation for the assessment of individual students with the aim of improving teaching and learning.

Since the late 1980s, researchers have begun to pay more and more attention to how FA can enhance learning, through both empirical studies and discussion papers, especially in UK and Australian contexts (cf., Andrade & Cizek, 2010; Black et al., 2003; Black & Wiliam, 1998a; Sadler, 1989; Torrance, 1993; Torrance & Pryor, 1998). This was generally in line with the traditional definitions given by Scriven and Bloom et al., but there was still a fine distinction. In the traditional definitions, researchers generally said FA was "for the purpose of improving" without indicating whether that was intention only or must also include resulting action. However, in the new definitions, the researchers implied that FA occurs only when there are beneficial

effects. For example, Sadler (1989, p.120) said FA "is concerned with how judgments about the quality of student responses (performances, pieces, or works) can be used <u>to shape and improve the student's competence</u> by short-circuiting the randomness and inefficiency of trial-and-error learning (emphasis added)". Black and Wiliam (1998c) said "assessment becomes 'formative assessment' when the evidence is actually <u>used to adapt the teaching work to meet the needs</u>" (p.2) (emphasis added). However, this absolute requirement for actual beneficial effects was relaxed a little in the latest definition of FA.

> Practice in the classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, <u>to make decisions</u> about the next steps in instruction <u>that are likely to be better</u>, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited. (Black &Wiliam, 2009, p. 9) (Emphasis added)

By using the terms—"to make decisions" and "are likely to be better", Black and Wiliam acknowledged both the intention for and the probabilistic nature of the beneficial effects of FA on student learning. Andrade (2010), in his summary chapter for the *Handbook of Formative Assessment*, used the word "purposes" to define FA, without specifying if FA referred to actual beneficial effects or just intended beneficial purposes.

> [A]ny definition of formative assessment must be grounded in its purposes, which include: (1) providing information about students' learning to teachers and administrators in order to guide them in designing instruction; and (2) providing feedback to students about their progress in order to help them determine how to close any gaps between their performance and the targeted learning goals. (Andrade, 2010, pp. 344-345)

From the above definition, it can be seen that feedback is an essential part of FA and the nature of feedback determines if an assessment is FA or not.

At the same time, the understanding of FA has been extended from referring to all the possible assessment methods used in classroom contexts to the processes involved in assessment (e.g., Black & Wiliam, 2009; Cowie & Bell, 1999; Gipps, 1994; Wiliam, 2010). An example of regarding assessment as methods was Black and Wiliam's (1998c) definition of FA: "all those activities undertaken by teachers, and by their

students in assessing themselves, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (p.2). However, Cowie and Bell (1999) defined FA as process, "the process used by teachers and students to recognize and respond to student learning in order to enhance that learning, during the learning" (p.101). Based on empirical data, they found three steps involved in each of the two types of FA they identified. Planned FA, which the teacher has planned in advance, involves the teacher eliciting, interpreting, and acting on the assessment information; and interactive FA, which arises out of classroom interactions, involves the teacher noticing, recognizing and responding to student thinking during student-teacher interaction. The two three-step processes correspond to the three steps mentioned in Cizek's (1997) definition: gathering, synthesizing, and use of information about student learning, and that in McMillan's (2007a, p. 8) definition: "collection, evaluation, and use of information to help teachers make decisions that improve student learning". This three-step process is also reiterated in Black and Wiliam's (2009, p. 7) recent definition of FA: "evidence about student achievement is elicited, interpreted, and used…".

Since the landmark review by Black and Wiliam (1998a), researchers in the general education field, especially the Assessment Reform Group in the UK, have been looking for features or principles or strategies that make an assessment practice formative (e.g., Black & Wiliam, 2003; Carless, 2011; Cizek, 2010; Shepard, 2006, 2008; Wiliam, 2010; Wiliam & Leahy, 2007; Wiliam & Thompson, 2008). The most recent and comprehensive synthesis is from Cizek who synthesized a list of ten characteristics of FA:

1   Requires students to take responsibility for their own learning.

2   Communicates clear, specific learning goals.

3   Focuses on goals that represent valuable educational outcomes with applicability beyond the learning context.

4   Identifies the student's current knowledge/skills and the necessary steps for reaching the desired goals.

5   Requires development of plans for attaining the desired goals.

6   Encourages students to self-monitor progress toward the learning goals.

7   Provides examples of learning goals including, when relevant, the specific grading criteria or rubrics that will be used to evaluate the student's work.

8   Provides frequent assessment, including peer and student self-assessment and assessment embedded within learning activities.

9 Includes feedback that is non-evaluative, specific, timely, related to the learning goals, and provides opportunities for the student to revise and improve work products and deepen understandings.

10   Promotes metacognition and reflection by students on their work. (Cizek, 2010, p.8)

It can be seen that an FA practice might have one or more of the features in the above list, but individual features alone cannot be used to identify whether an assessment is FA. For example, "communicates clear, specific learning goals", the second strategy in the above list, will not ensure that an assessment is formative. As Cizek pointed out, "not all of the characteristics must be present for an assessment to be considered formative", and "not all of the elements … have to be combined in any particular application of formative assessment" (Cizek, 2010, pp. 7-8). However, he did not specify what are the essential features to identify FA. Therefore, while this list of features may be useful for teachers to develop their own FA practices, it is difficult for a researcher to use this list to identify FA practices.

In essence, the above discussion indicates that FA should have two key features. First, it should fulfil the requirements of assessment, that is, it should collect information, interpret the information, and then make use of the interpretation (e.g., Cizek, 1997; Cowie & Bell, 1999; McMillan, 2007b). Second, it should be assessment used to serve the purposes of improving teaching and learning. So far, the second feature has received more emphasis than the first feature, as Andrade stressed: FA "is not a particular tool but rather a matter of the uses to which assessment data are put. … formative assessment refers to the purposes of assessment information, not to particular assessment procedures or instruments" (Andrade, 2010, p. 344).

Assessment for learning (AfL)

Assessment for Learning (AfL) originated in the UK in early 1990s (Popham, 2011) and gained its popularity mainly through the work of the Assessment Reform Group (ARG) (ARG, 1999, 2002; Black et al., 2003; Black, Harrison, Lee, Marshall, & Wiliam, 2004; Gardner, 2006b; Stobart, 2008; Wiliam, 2009). This term was chosen by the ARG mainly to avoid possible misunderstandings of FA and to emphasize the true nature of FA (ARG, 1999; Stobart, 2008). In the 1999 pamphlet, the authors pointed out that the term "formative" is open to a variety of interpretations and often means no more than that assessment is carried out frequently and is planned at the same time as teaching, which does not necessarily help learning and therefore should not be called formative. Stobart (2008) also pointed out that regarding regular classroom tests which are used for monitoring progress as formative is a misunderstanding of FA. Therefore, AfL is used to emphasize the true nature of FA, which is not about when to conduct assessment but about the function of AfL, which is to be conducive to learning.

In essence, AfL is the same as FA, as shown in the widely used definition of AfL, provided by ARG (2002, pp. 2-3): AfL means "the process of seeking and interpreting evidence for use by learners and their teachers, to identify where the learners are in their learning, where they need to go and how best to get there".

It can be seen that AfL, like FA, acknowledges that assessment is a process involving three steps—seeking, interpreting, and using, and it focuses on the purposes of assessment for improving student learning. Just as FA is contrasted with SA, AfL is also contrasted with assessment of learning (AoL) (ARG, 1999). While AfL is about how to use assessment to enhance student learning, AoL is mainly to summarize what students have learned and report assessment outcomes. Therefore, some researchers use AfL and FA interchangeably, and use AoL and SA interchangeably (e.g., Harlen, 2007).

Learning-oriented assessment (LOA)

LOA was coined and used by a group of researchers headed by David Carless situated in Hong Kong in recent years (Carless, 2007; Carless, Joughin, Liu, et al., 2006; Carless, Joughin, & Mok, 2006; Keppell & Carless, 2006). According to Carless (2007), this term was promoted to "avoid the confusion and doubts about" FA (p.58). By using LOA, the researchers emphasized the learning purpose of assessment while acknowledging that the same assessment might also perform a certification purpose. It can be seen that this term is essentially the same as AfL, both of which are primarily concerned with the formative functions of assessment.

Alternative assessment (AA)

Alternative assessment (AA) started to attract attention around the beginning of the 1990s when new forms of assessment other than the traditional testing format were suggested and tried out as a way to reflect the new cognitive constructivist approach to teaching and learning and to meet the needs of assessment reform (Anderson, 1998; Marzano, Pickering, & McTighe, 1993; Shepard, 1989, 2000). With it came a few other similar terms such as performance assessment and authentic assessment (Marzano et al., 1993; Wiggins, 1989). Although sometimes they are used interchangeably, there are some fine distinctions, as Marzano et al.(1993) explained:

> *alternative assessment* applies to any and all assessments that differ from the multiple-choice, timed, one-shot approaches that characterize most standardized and many classroom assessments. … *authentic assessment*, … conveys the idea that assessments should engage students in applying knowledge and skills in the same way they are used in the 'real world' outside school. … *performance assessment* refers to a variety of tasks and situations in which students are given opportunities to *demonstrate* their understanding and to thoughtfully *apply* knowledge, skills, and habits of mind in a variety of contexts. (italicized original) (p.13)

It can be seen that alternative assessment is an umbrella term encompassing all the possible assessment instruments/procedures other than the traditional "single-event, discrete-point, multiple-choice tests that result in numerical scores and the ranking of individuals" (Fox, 2008, p. 97). Therefore, alternative assessment covers both

performance assessment and authentic assessment, as well as task-based assessment, dynamic assessment, etc. (Fox, 2008).

In sum, unlike CA and TBA that focus on classroom contexts, or FA/AfL/LOA that focus on assessment purposes, alternative assessment focuses on the forms/methods of assessment. Because alternative assessment includes a wide range of possible assessment methods, researchers sometimes examine one type such as portfolios and then use it interchangeably with alternative assessment. Although alternative assessment embodies constructivist and later social-cultural theories of learning and emphasizes the formative function of assessment (Anderson, 1998; Fox, 2008), its actual purpose/function is determined by how it is used rather than how it is designed (Fox, 2008).

Dynamic assessment (DA)

DA is often traced back to Vygotsky's work that is concerned with the development of a person's potential abilities (Grigorenko, 2009; Leung, 2007; Poehner, 2007, 2008, 2009; Poehner & Lantolf, 2005; Sternberg & Grigorenko, 2002). It is built upon the assumptions that "conventional (also referred to as unassisted or static) assessment might not adequately capture the level of cognitive development" (Grigorenko, 2009, p. 113); and a person's abilities are "malleable and flexible rather than fixed" (Sternberg & Grigorenko, 2002, p. 1). Vygotsky's concept of Zone of Proximal Development (ZPD) "provides the theoretical framing and operational space for DA" (Leung, 2007, p. 258). Although DA has a long history, "it has not sparked any widespread pedagogical revolutions, even in the domain of special education where so much DA work has been conducted" (Poehner, 2008, p. 175). In recent years, the language assessment field has seen a rising interest in DA (e.g., Leung, 2007; Poehner, 2007, 2008, 2009; Poehner & Lantolf, 2005).

DA is an "umbrella term used to describe a heterogeneous range of approaches" in psychology and education, and the core of DA is in blending instruction into assessment (Elliott, 2003, p. 16). It is characterized by integrating mediation/intervention into the assessment procedures in order to understand and promote learner development (Elliott, 2003; Grigorenko, 2009; Leung, 2007; Poehner,

2007, 2008, 2009; Poehner & Lantolf, 2005). According to how mediation should be offered, there have been two prevailing models of DA: interventionist and interactionist. Interventionist DA emphasizes standardized mediation offered to learners, in order "to maximize the assessment's objectivity" (Poehner, 2008, p. 45). In interactionist DA, constantly fine-tuning mediation is offered to co-construct a ZPD with the learner in order to optimally promote learner development.

To make the original one-on-one model of DA more applicable to classroom contexts, Poehner (2009) proposed a group model of DA (G-DA) for the L2 classrooms, drawing on Petrovsky's (1985) perspectives on individual and group abilities that have emerged in psychology. He proposed two approaches to G-DA: concurrent G-DA, "the teacher dialogues with the entire group, …the interaction shifts rapidly between primary and secondary interactants as one learner's question, struggle, or comment sets the stage for another's contribution", and cumulative G-DA, "the teacher conducts a series of one-on-one DA interactions, …the understanding that each subsequent one-on-one exchange will have the advantage of building on earlier interactions that the class witnessed" (Poehner, 2009, p. 478). Poehner argued that, although G-DA is very similar to teacher-led class discussion, the two are different in that G-DA is informed by theory and a teacher-led class discussion sometimes may be an intuitive practice and the teacher may not be very clear why a discussion is successful or unsuccessful.

It can be seen that the distinctive feature of DA is its foundation on a theory of learning and cognitive development, whereas the other concepts discussed above lack this kind of explicit theoretical basis. Guided by the socio-cultural theory of learning, DA, or G-DA, distinguishes itself from other concepts especially in terms of administration procedures, namely, during DA there is mediation/intervention (in a dynamic way) during the assessment process.

2.2.2 CA defined in this study

The above discussion showed that CA contains several key features: it is a process involving three key steps, it is conducted in classroom contexts, it may serve multiple purposes, it may take various forms, and there may be mediations during an

assessment process. Because the present study intended to describe and understand teachers' CA practices in their own specific classrooms, the author chose the term CA for its emphasis on classroom contexts and allowance for variability in other dimensions of CA.

Therefore, in this study, CA refers to a process occurring in the classroom context, of collecting information about student learning, making a judgment about the information collected, and then making use of the judgment made. While the process of eliciting-interpreting-using is at the very core of CA, a teacher may conduct each of the three steps in a variety of ways. For the evidence-collecting step, a teacher may use various methods from traditional multiple-choice tests to alternative methods like portfolios, or by providing online feedback to scaffold student learning (e.g., Cowie & Bell, 1999; Shepard, 2005). For the judgment-making step, a teacher may interpret a student's performance in relation to the whole group who also take the assessment (NRM) or against some pre-specified learning objectives (CRM). Besides, a teacher may do the judgment all by him-/herself, or involve students in doing self- / peer-assessment. For the judgment-using step, a teacher may use the judgment achieved either for summative functions (SA) such as grading and reporting or for formative functions (FA) such as improving teaching and learning. Furthermore, CA may occur any time during an instructional program, although SA tends to occur at the end of an instructional program or at key points during the program.

One point needs to be emphasized. The above-mentioned possibilities for each of the three steps are not restrictive. Instead, they are just guidelines for the researcher to explore the CA practices in the selected teachers' actual contexts. During the study, the researcher tried her best to remain open and sensitive to all other possibilities.

2.3 Models / frameworks of CA

Within the past decade, attempts have been made to theorize CA or FA/AfL, both in the general education field (Black & Wiliam, 2006b; Brookhart, 1997; Brookhart, Walsh, & Zientarski, 2006; Cowie & Bell, 1999; Harlen, 2007; Wiliam, 2010; Wiliam & Thompson, 2008), and in language assessment (Davison & Leung, 2009; Hill & McNamara, 2012). According to the emphasis of each, these models can be roughly

divided into three groups: descriptive, prescriptive, or research-oriented ones. In this section, existing models/frameworks are described and critiqued.

2.3.1 Descriptive models

So far, there are four descriptive models or frameworks for FA or CA. Cowie and Bell's model (1999) focuses on FA processes during classroom teaching. The two frameworks developed by Harlen (2007) emphasize the components and variables of assessment as a system as well as the relationship between FA and SA. Brookhart's model (Brookhart, 1997; Brookhart et al., 2006) emphasizes the causal relationships between CA and student motivation and learning. All of these models are within the general education field.

Cowie and Bell's model

Cowie and Bell's model (1999) concerns the nature of the FA process. It was derived from a two-year project situated in the science classrooms of ten teachers in Years 7-10 in New Zealand. Aiming to reveal the process of FA in these classrooms, the researchers conducted 128 classroom observations and found two kinds of FA: planned FA and interactive FA (Figure 2.1).
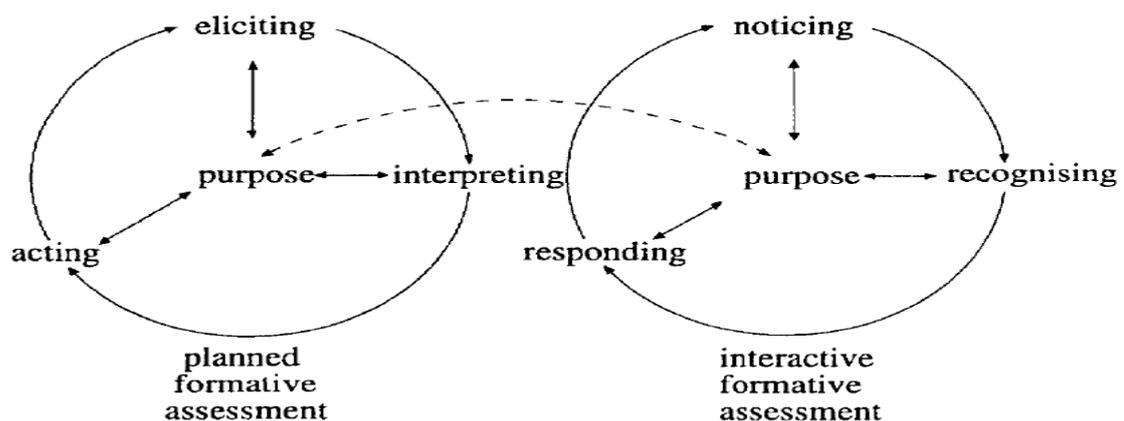
Figure 2.1 Cowie and Bell's Model of FA (source: Cowie & Bell, 1999, p.113)

According to Cowie and Bell, if an episode of classroom teaching and learning contains three segments (a teacher's assigning a pre-planned task to elicit student

performances, making an interpretation of such performances, and then acting upon such data collected), then the teacher has conducted a planned FA event. If, during the teacher-student or student-student interactions, an episode occurs comprising the following three parts (a teacher noticing some performance from some student, recognizing the student's problems/difficulties as reflected through such performances, and then responding to such problems/difficulties), then the teacher has conducted an interactive FA event. Generally speaking, planned FA tended to be prepared in advance and used with the whole class, but interactive FA was contingent upon spontaneous classroom interactions and used with individual students or small groups. These two kinds of FA were linked through the purposes of FA, because a teacher's planned FA purpose might be modified based on his/her judgment of student performances and consequently the teacher might conduct an interactive FA event.

Because it pointed out the essential features of FA processes within the context of regular classroom activities, this framework was significant for the present study in terms of helping identifying those FA practices that were highly embedded in everyday classroom teaching and learning and those that might not be recognized by the participant teachers in the Chinese EFL context. However, since this framework was developed in a very different educational context, the researcher remained cautious and sensitive during the study to see if the collected data could match the framework.

Harlen's models

Based on her previous work on the relationship between FA and SA (Harlen, 2005, 2006; Harlen & James, 1997), Harlen (2007) proposed two models of CA. The first one focused on the components of CA as a system (Figure 2.2).

```
                    ┌ Purpose        ➤ Formative
                    │                ➤ Summative
                    │
                    │                ➤ Helping learning
                    │                ➤ Internal school tracking of student progress
                    │  Use           ➤ Informing next teacher, parents, students
                    │                ➤ Certification/accreditation of individual learning
                    │                ➤ Selection
                    │                ➤ Monitoring school performance
                    │                ➤ Monitoring national standards
                    │
                    │                ➤ Regular work
                    │                ➤ Embedded tasks created by teachers
                    │  Type of task  ➤ Embedded tasks externally provided
                    │                ➤ Tests/tasks created by teachers
                    │                ➤ Tests/tasks selected from external pool or bank
                    │                ➤ Tests/tasks prescribed externally
Variables in        │
an            ──────┤  Agent of      ➤ Teacher
assessment          │  judgment      ➤ Teacher and students together
system              │                ➤ External agent
                    │
                    │  Basis of      ➤ Norm-referenced
                    │  judgment      ➤ Criterion-referenced – detailed, task-specific
                    │                ➤ Criterion-referenced – general, not task-specific
                    │                ➤ Student-referenced (ipsative)
                    │
                    │  Form of       ➤ Mark or score
                    │  report or     ➤ Profile of achievement against specified criteria
                    │  feedback      ➤ Overall level or grade achieved
                    │                ➤ Rank order
                    │                ➤ Comment or oral feedback
                    │
                    └  Moderation    ➤ Specified within-school moderation procedures
                                     ➤ Specified inter-school moderation procedures
                                     ➤ External inspection of procedures/evidence
                                     ➤ Informal – left to school
                                     ➤ Teacher/school/centre accreditation
```

Figure 2.2 Harlen's Model of CA: The Component Dimension (source: Harlen, 2007, p.15)

This model shows that assessment is a system incorporating seven components, each of which further contains a number of variables. Harlen further pointed out that these components and variables can serve as an analytic tool to describe assessment procedures and systems, including both FA and SA (Harlen, 2007, p.14). It can be seen that these components cover why teachers assess (purpose), how teachers assess

25

(type of task, agent of judgment, basis of judgment, type of report or feedback), use of assessment results (use), and quality assurance procedures (moderation).

The second model, which focuses on the time dimension of CA, suggests an ideal way in which evidence collected for formative purposes could be used for summative purposes as well (Figure 2.3).
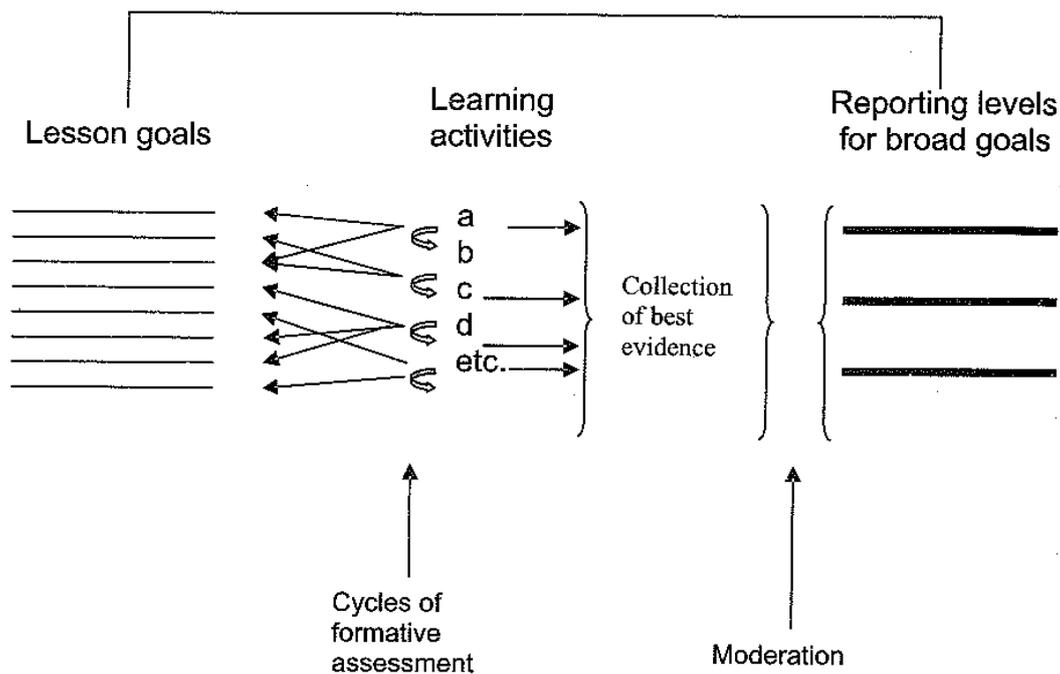
Figure 2.3 Harlen's Model of CA: The Time Dimension (source: Harlen, 2007, p.128)

According to this model, the origin of evidence for both formative and summative purposes is the series of relevant learning activities throughout a course. Around each learning activity is a cycle of FA. Cycles of FA use the evidence of students' performance on these learning activities to help learning towards specific lesson goals. Such cycles will run through the whole course. When it is time for the summative report, a collection of best evidence will be selected to reflect students' attainment in terms of course goals. So specific lesson goals are a kind of breakdown of the broader course goals. This model puts CA into the time frame of a course and reflects the nested relationships between CA at different levels: at the activity level and at the course level.

While the component dimension of CA can be used to generate cross-sectional pictures of CA practices, the time dimension of CA points out an ideal relationship between FA and SA in a longitudinal sense. Harlen's two models indicate that an investigation of CA practices should not only include descriptions of specific CA practices but also the relationship of different CA practices over time. Therefore, in this study, to reveal the comprehensiveness and complexity of the CA practices in the selected oral English classes, the researcher examined both the component dimension and the time dimension of CA practices.

Brookhart's model

Unlike the above three models which deal with teachers' CA practices, Brookhart's model (Brookhart, 1997; Brookhart et al., 2006) focuses on the causal relationship between CA practices, student motivation, and student learning (Figure 2.4).
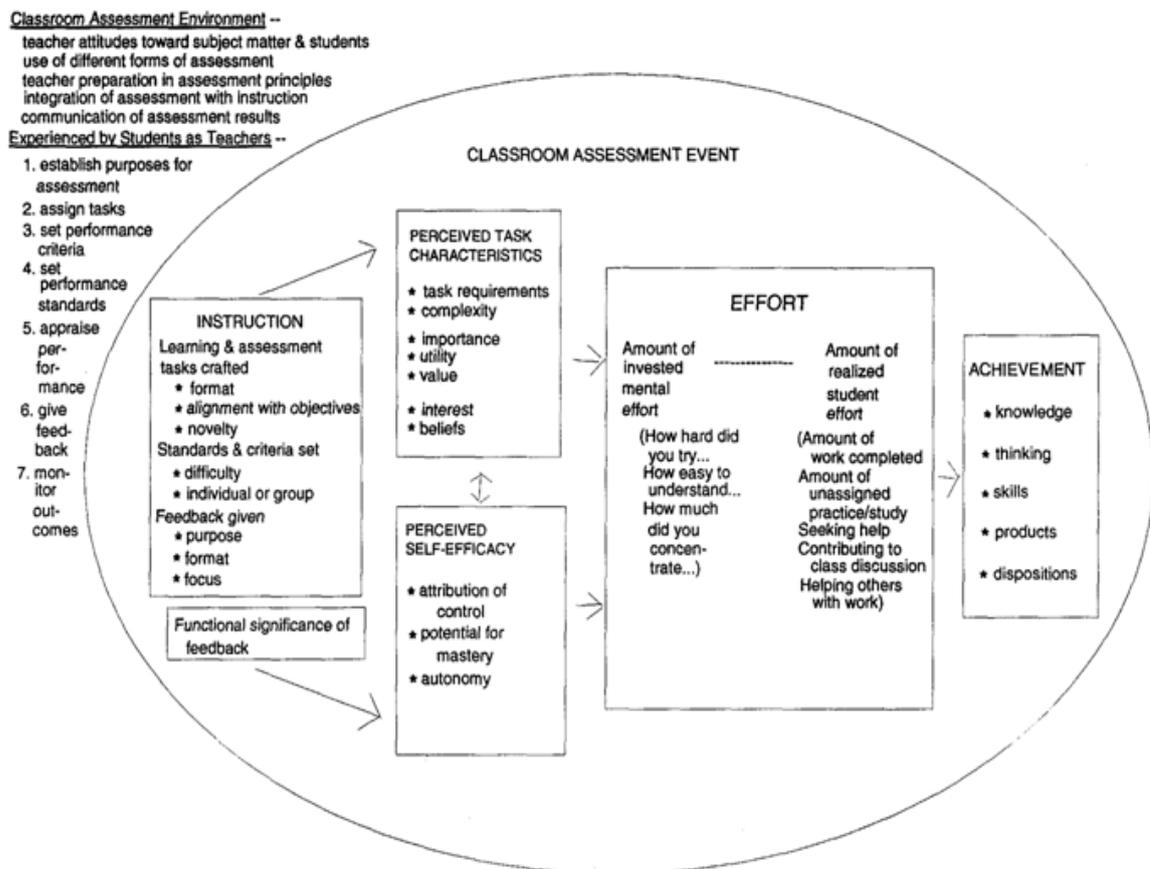
Figure 2.4 Brookhart's Model of CA (source: Brookhart, 1997, p.164).

Based on a synthesis of two different bodies of literature (the classroom assessment environment and social-cognitive theories of learning and motivation), Brookhart proposed this model to indicate that certain characteristics of assessment tasks such as task novelty or difficulty may influence how students perceive such tasks and their perceived self-efficacy, which in turn may influence their effort and then their achievement.

Later, through a series of empirical studies to test this model (Brookhart & DeVoge, 1999; Brookhart & DeVoge, 2000; Brookhart & Durkin, 2003) and further literature review (Brookhart, 2004), Brookhart revised the model (Figure 2.5).
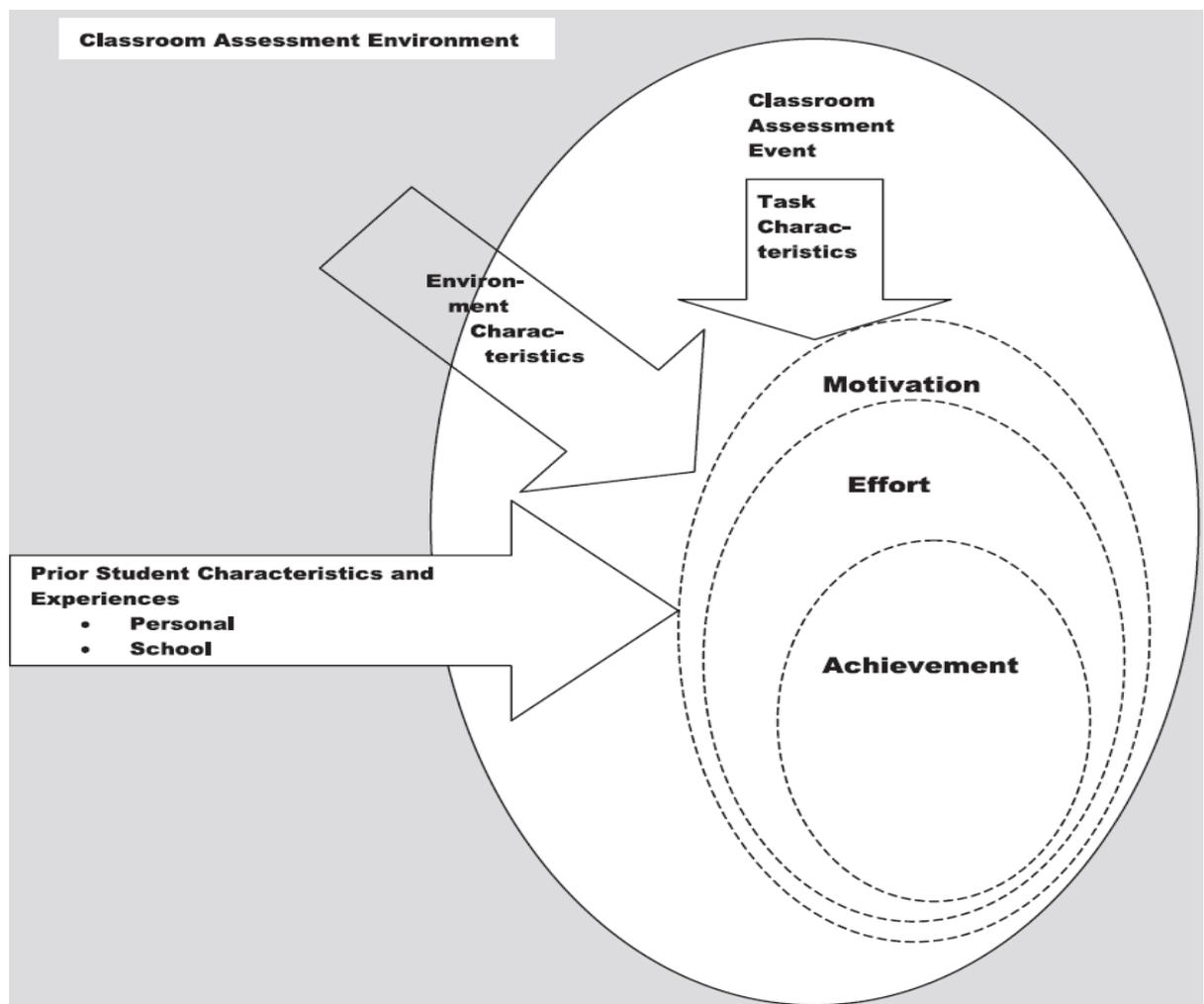
Figure 2.5 Brookhart's Revised Model of CA (Source: Brookhart et al., 2006, p.153).

In the revised model, she retained the causal relationship between students' perceptions of assessment tasks, student effort, and student achievement. However, she made two major changes. First, she inserted one factor—student motivation—before student effort in the causal chain, because, in her opinion, motivation (wanting to do something) is different from effort (actually doing it) (Brookhart et al., 2006, p. 152). Second, the concept *classroom assessment event*, which was the focus of the 1997 model, was referred to as small contexts in the 2006 model. In this revised model, she added the factor of *classroom assessment environment*, referred to as large contexts. While *classroom assessment events* often take the form of classroom activities, *classroom assessment environment* is operationalized as teacher assessment practices in a classroom that "make[s] the assessment experience different in different classrooms" (ibid, p.152).

Although Brookhart's model does not specify the elements involved in teachers' assessment practices, as Harlen's and Cowie and Bell's models do, its strength lies in proposing the causal relationship between teachers' CA practices and the impacts on student motivation and learning as perceived by students. Therefore, Brookhart's model was used to develop student interview questions and the student questionnaires for this study to find out how CA practices would affect students' motivation, effort, and their sense of achievement in their oral English classes (cf. Section 3.4).

2.3.2 Prescriptive models

Two prescriptive models were found (Wiliam & Thompson, 2008; Davison, 2008, cited in Davison & Leung, 2009) which aimed to instruct teachers on how to conduct their CA so as to make it beneficial for student learning.

Wiliam and Thompson's model

Out of concern for teacher professional development so as to improve education, Wiliam and Thompson (2008) put forward a framework of FA to show how FA can be implemented to facilitate student learning (Figure 2.6). This framework reappeared in

Wiliam's 2010 article.

	Where the learner is going	Where the learner is right now	How to get there
Teacher	Clarifying learning intentions and sharing and criteria for success (1)	Engineering effective classroom discussions. activities and tasks that elicit evidence of learning (2)	Providing feedback that moves learners forward (3)
Peer	Understanding and sharing learning intentions and criteria for success (1)	Activating students as instructional resources for one another (4)	
Learner	Understanding learning intentions and criteria for success (1)	Activating students as the owners of their own learning (5)	

Figure 2.6 Wiliam and Thompson's Model of FA (source: Wiliam & Thompson, 2008, p.63; Wiliam, 2010, p.31)

Starting from the three key purposes of FA/AfL(as shown in the first row of the above figure), which were derived from Ramaprasad's conceptualization of feedback (Ramaprasad, 1983, p. 4), and which were also regarded as the three purposes of instructional processes (Wiliam, 2010, p.30), the authors used this framework to reveal how the three key purposes can be achieved through five FA strategies, which are numbered in the above figure. In this way, FA and instruction are integrated.

Davison's model

In a similar vein, Davison's model (2008, cited in Davison & Leung, 2009) (Figure 2.7) shows the key steps involved in TBA so as to bring about desirable learning outcomes. It is like a step-by-step guideline for teachers to conduct effective TBA in their own classrooms.

FIGURE 1
A Framework for Teacher-Based Assessment (Davison, 2008)
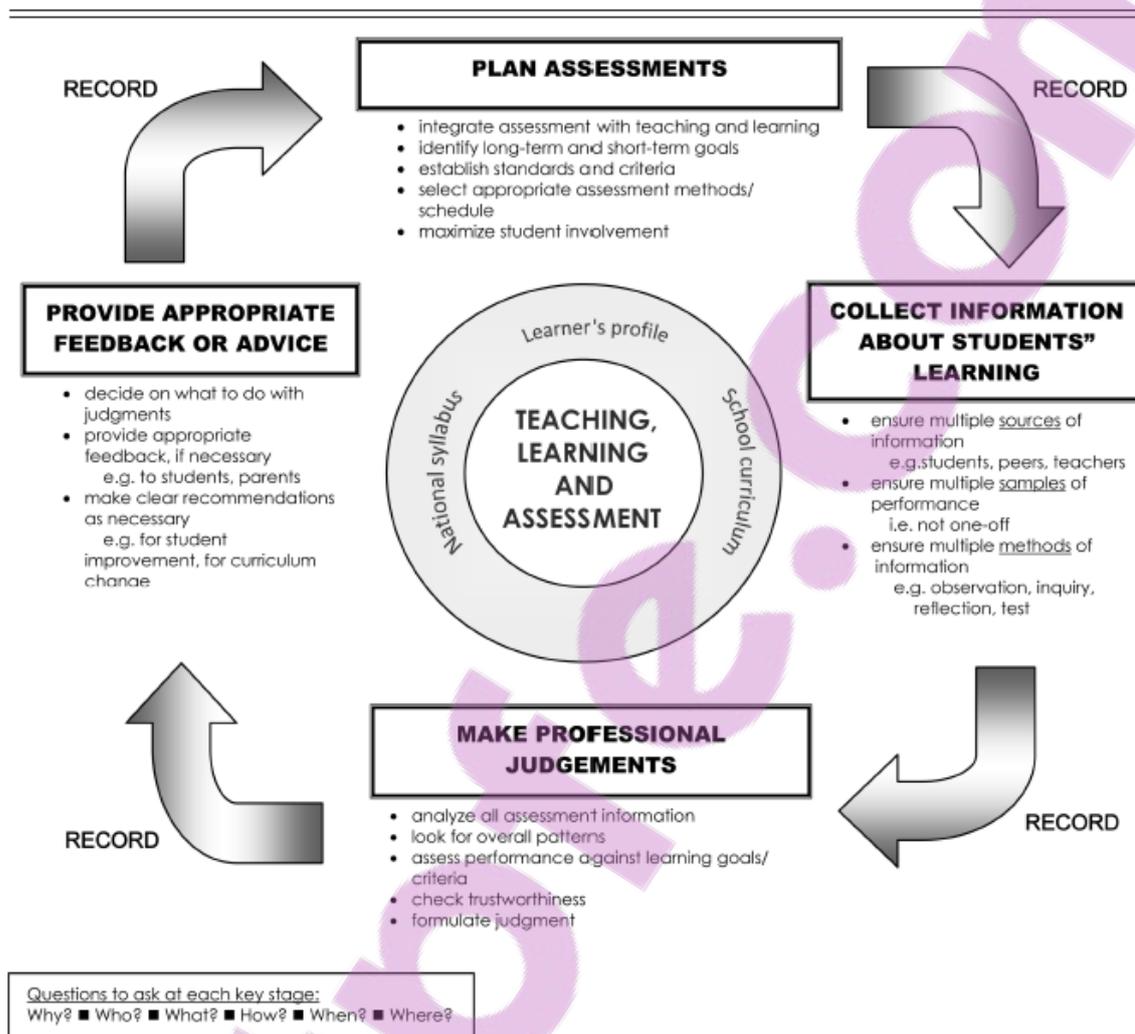
Figure 2.7 Davison's Model of TBA (source: Davison & Leung, 2009, p.395)

It can be seen that Wiliam and Thompson's model is more instructionally oriented while Davison's model is more assessment-process oriented. Both models reveal the key features and ideal procedures of effective FA/AfL or TBA, but neither is suitable to serve as a descriptive tool to display teachers' CA practices. Instead, both may serve as a frame of reference for the researcher of the present study to examine the effectiveness of the three teachers' CA practices investigated in the present study.

2.3.3 Research-oriented models

So far, two models have been put forward to guide research on FA.

Black and Wiliam's model

Based on the data obtained from one project aiming to enhance learning through FA (cf. Black & Wiliam, 2006a), Black and Wiliam (2006b) categorized the changes identified in the participant teachers' FA practices into four aspects and argued that these four aspects "provide the minimal elements of a theory of formative assessment" (p.84). The four aspects are:

"Teachers, learners, and the subject discipline", which concerns the interplay between "teachers' views of the nature of the subject matter", "the selection and articulation of goals", and their "models of cognition and of learning" (p.85)

"The teacher's role and the regulation of learning," which concerns to what extent the way teachers plan and set up a lesson can bring about "teachable moments" (pp. 86-87)

"Feedback and the student-teacher interaction," which concerns the determinants of effective feedback and teachers' ability to adapt to the different ZPDs in a class so as to both bring about "enhanced student-teacher interaction" and "handle differentiation" successfully (pp. 87-91)

"The student's role in learning," which concerns the interplay between students "own beliefs and implicit models of learning" and their achievement (pp.91-94)

The authors also found it productive "to think of the subject classroom as an 'activity system'" (p.83) and found that these four aspects of changes fall into the top part of the Activity Theory: tools-subjects-objects (Engeström, 1987), as can be seen in Figures 2.8 and 2.9.
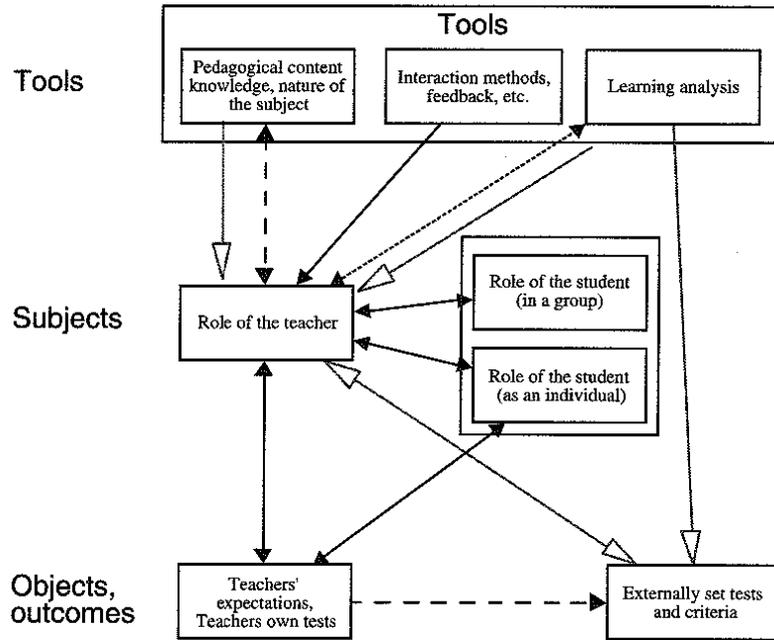
Figure 2.8 Black and Wiliam's Application of the Activity Theory in Understanding FA (Source: Black & Wiliam, 2006b, p.95)
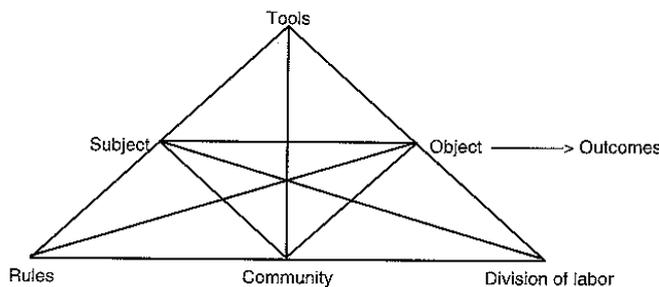
Figure 2.9 Engestrom's Canonical Representation of the Activity Theory (Source: Black & Wiliam, 2006b, p.96)

Obviously, this model regards FA as one of the tools in the activity system of a classroom, and it emphasizes "the way that tools (a term that includes procedures such as feedback and peer assessment) alter teacher-student interaction, the role of the teacher in the regulation of student learning, and students' own role in their learning, and how such changes bear upon the outcomes of learning work" (Gardner, 2006a, p. 4). It can be seen that this framework puts FA within the activity theory framework (cf., Engeström, 1987, 1999, 2001), a sub-field of socio-cultural theory (cf., Lantolf,

2006; Lantolf, 2000), which to some extent indicates the applicability of Activity Theory to FA research.

However, this particular feature makes this framework not as informative as the other models when it comes to understanding and analyzing CA, because FA is not fore-grounded in this framework but immersed within the complexity of classroom contexts. For example, while Black and Wiliam point out the importance of investigating the teacher's role and the student's role in classroom teaching and learning as well as the teacher's feedback, Harlen's component-dimension model (Figure 2.2) makes these aspects more clearly related to assessment, as shown through these components: "agent of judgment" and "form of report or feedback". While Black and Wiliam suggest careful investigation of teacher-student interaction, Cowie and Bell's model (Figure 2.1) reveals the typical patterns that can be considered as FA. In addition, this framework was only for FA, and SA was not included. Considering the focus of the present study was on teachers' CA practices, including both FA and SA, the author found this framework not as informative as the other models, though it did suggest the socio-cultural approach and the important relationships and aspects that are worth investigating when researching FA.

Hill and McNamara's framework

Based on an empirical study of two Australian school classrooms where students aged 11 to 13 were studying Indonesian as a foreign language, Hill and McNamara (2012) proposed a framework for researching classroom-based assessment processes (Table 2.1).

Table 2.1 Hill and McNamara's framework on CA (Source: Hill & McNamara, 2012, p.415).

1. What do language teachers do?	
1.1. Planning Assessment	Is there planning for assessment? How detailed is planning? What is its intended relationship to instruction? How does it relate to external standards and frameworks?
1.2. Framing Assessment	Is assessment made explicit to learners? How is this done?
1.3. Conducting Assessment	What opportunities does the classroom provide for assessment? Does assessment tend to focus on the class, group/pairs of students or individuals?
1.4. Using Assessment Data	How is assessment-related information used? Teaching
	Learning (feedback) Person-referenced
	Task-referenced
	• Confirmatory
	• Explanatory
	• Corrective
	Reporting
	Management
	Socialization
2. What do teachers look for?	
What information about valued enterprises, qualities and standards is available?	
2.1. In Advance	in written/verbal instructions and/or assessment rubrics?
2.2. In Feedback	in written and/or verbal feedback?
2.3. In Reporting	in reporting deliberations and/or in written reports?
3. What theory or 'standards' do they use?	
3.1 Teacher Theories & Beliefs	What does the data reveal about teachers' beliefs about • the subject or content area • second language learning and teaching, and • the nature of assessment?
4. Do learners share the same understandings?	
4.1. Learner Theories & Beliefs	What does the data reveal about learners' beliefs about • second language learning, and • the nature of assessment?

This framework is really comprehensive. For one thing, it covers many aspects of CA, from the preparation of assessment (1.1, 1.2) to the assessment process (1.3, 1.4) to assessment constructs (2) to the epistemological basis for such practices (3) and finally to students' understanding of assessment (4). In other words, almost all the relevant aspects of CA in a classroom context are included in the framework. For another, it specifies the sources where a researcher should look for relevant information or the research questions that a researcher should ask. Moreover, it can be applied to both FA and SA. Therefore, this framework sets a research agenda for those who want to investigate CA thoroughly.

However, by the time this article was published, the data collection for the present study had been finished. Though the present study design was not informed by this

comprehensive framework, this framework contributed to the data analysis of the present study (cf. Section 3.4).

2.3.4 Summary

Two features are revealed from the above discussion on the existing models / frameworks of CA/FA.

First, except for Black and Wiliam's (2006b) framework, which sets out to establish an association with Activity Theory, none of the other models / frameworks demonstrate explicit links with any learning theories.

Second, these models have approached CA/FA from different perspectives: from the perspective of its internal components and mechanism (Cowie & Bell's model, Harlen's component-dimension model, Davison's model), from its functional perspective (Brookhart's model, Harlen's component-dimension model, Wiliam& Thompson's model), or from its temporal perspective (Harlen's time-dimension model). This shows that CA is a multi-faceted, multi-dimensional, and multi-functional tool in classroom teaching and learning.

While so much progress has been achieved concerning theorizing CA, few existing empirical studies on CA in L2 assessment field have been informed by the above-mentioned models, as can be seen from the following review.

2.4 Research on CA in L2 assessment field

While there is an extensive literature on CA in the general education field (cf., Andrade & Cizek, 2010; Black & Wiliam, 1998a; Brookhart, 2004; Crooks, 1988; Gardner, 2006b; Harlen, 2007; McMillan, 2007c; Natriello, 1987; Phye, 1997), the following part of the literature review will concentrate on studies on CA in L2 assessment field, because the present study was situated in an EFL context. Since CA is contextually grounded (Carless, 2011; Leung & Mohan, 2004) and "assessment takes different shapes depending on subject matter (including the formal and informal curricula) and developmental level" (Calfee & Masuda, 1997, p. 69), the existing

research findings about CA in similar contexts or in a similar subject would very well inform the present study.

An examination of these studies revealed that CA research has received increasing attention over the past decade. Specifically, research on CA has extended from EAL (English as an additional language) and ESL (English as a second language) contexts to EFL (English as a foreign language) contexts, and advances have been made mainly in the following three areas: FA-SA relationship, characteristics of teachers' CA practices, and relationship between pre-existing or newly-designed CA practices and student learning. These three groups of studies will be reviewed in detail below.

2.4.1 FA-SA relationship

The FA-SA relationship was mainly investigated by researchers concerned with the construct of CA/FA in primary-school EAL contexts in Britain (Leung & Mohan, 2004; Rea-Dickins, 2001, 2006, 2007b; Rea-Dickins & Gardner, 2000).

Through 28 teachers' self-report of their CA practices and what each practice was used, corroborated with some classroom observation data, Rea-Dickins and Gardner (2000) found that the same assessment procedures could serve both formative and summative purposes; thus, the distinction between FA and SA is not straightforward. By examining three assessment activities, each of which was conducted by one of three participating teachers, Rea-Dickins's 2001 paper summarized three identities of CA: bureaucratic identity, which is mainly concerned with accountability issues; pedagogic identity, which is mainly concerned with providing feedback for teaching; and learning identity, which is mainly concerned with supporting and enhancing student learning. She found that one assessment practice embodied both the bureaucratic identity and the learning identity, which indicates that the distinction between FA and SA is not that clear-cut, echoing the finding from the previous study.

Unlike Rea-Dickins's two studies mentioned above, where those assessment activities under investigation were all identified by the participating teachers, in Leung and Mohan's (2004) study the assessment episodes examined were selected by the researchers according to the definition of AfL (ARG, 2002). In this study, by using

Vygotsky's ZPD theory and Halliday's systemic functional linguistics, the researchers found differences between AfL and formal AoL in all the four stages of assessment—(A) task setting, (B) student performance, (C) teacher assessment of task outcome, and (D) formative guidance. Their results revealed that the instruction-embedded FA practices took different forms from formal SA, especially at stage D when the teachers "guided students to reject incorrect answers and accept correct answers on the basis of reasons" (Leung & Mohan, 2004, p. 355). In other words, AfL was enacted through scaffolding. The distinctions found between AfL and AoL implied a clear-cut line between the two.

Taking Leung and Mohan's (2004) study findings into consideration, Rea-Dickins (2006) reconceptualised FA/CA as along a continuum from more planned to more unplanned and spontaneous, and focused on learners' engagement with some assessment episodes, not all of which were regarded by the teachers as assessment. One finding was that for one explicitly labeled SA episode, one student was taking it as a learning opportunity. In other words, this SA episode performed formative functions for this student's learning. Therefore, this study, from the perspective of learner engagement, found evidence that the same assessment procedure can serve both formative and summative functions, echoing the finding from her previous studies.

Rea-Dickins's (2007b) paper focused on teachers' decision-making when they planned their CA. Through classroom observation and teacher interview, she found evidence that sometimes teachers proposed to use assessment data for both summative and formative purposes, thus corroborating the findings from her previous studies.

In general, Rea-Dickins's studies demonstrated that the line between FA and SA may not be clear-cut. Specifically, teachers may plan an assessment event to serve both summative and formative purposes, an assessment event may actually serve dual purposes at the same time, and an SA may perform formative functions as well for some students. However, it should be noted that the above-mentioned studies were all situated in primary-school EAL/ESL contexts in Britain, where the teachers were all native English speakers working with EAL/ESL learners in mainstream courses

instead of language-learning programs. Therefore, to what extent their findings can be applied to tertiary-level EFL language-learning contexts deserves further investigation.

2.4.2 Characteristics of teachers' CA practices

Another group of studies mainly focuses on various aspects of teachers' CA practices, including assessment purposes, assessment methods, judgment-making processes, grading and reporting practices, and follow-up strategies. These studies can be further divided into two groups: general survey studies (Cheng, Rogers, et al., 2004; Cheng et al., 2008; Cheng & Wang, 2007) and studies focusing on one specific aspect of CA (Butler, 2009; Davison, 2004; Yin, 2010).

Cheng and her colleagues conducted two comparative studies, the first one a questionnaire survey (Cheng, Rogers, et al., 2004) and the second one an interview study (Cheng et al., 2008; Cheng & Wang, 2007), to reveal contextual differences among the classroom assessment practices of three groups of tertiary-level ESL/EFL teachers in Canada, Hong Kong, and Beijing. Their study instruments, the questionnaire and the interview guide, which were derived from government documents on good educational assessment practices (American Federation of Teachers, 1990; Joint Advisory Committee, 1993), cover a wide range of teachers' CA practices including assessment methods, purposes, procedures, grading, reporting, and follow-up strategies.

Regarding assessment methods, Cheng, Rogers, and Hu (2004) investigated the way the three groups of teachers assessed students' reading, writing, and speaking/listening respectively. Regarding the assessment of students' speaking/listening skills, the researchers classified assessment methods into three groups: instructor-made assessment methods such as "take notes" and "prepare summaries of what is heard", student-conducted assessment methods such as "oral presentation" and "oral discussion with other students", and non-instructor developed assessment methods such as "standardized speaking test" and "standardized listening test (p.370). Their results showed that, while oral presentation is the most common format and there was no statistically significant difference among the three groups concerning the use of oral presentation method, Beijing and Canada teachers used significantly more of the

following student-conducted assessment methods than Hong Kong teachers: oral interviews/ dialogues, oral discussion with each student, retelling a story, providing an oral description, oral reading / dictation, and giving oral directions. However, Beijing teachers used significantly less peer assessment than Canada and Hong Kong teachers.

In their interview study, Cheng and her colleagues (Cheng et al., 2008) did not investigate the assessment of particular skills. Instead, they classified assessment methods into two major groups: selection-response and supplied or constructed response methods/ formats. Their results showed that, regarding selection types of assessment methods, Beijing and Canada teachers more frequently used selection items than Hong Kong teachers, and Beijing teachers used standardized tests more frequently than Canada and Hong Kong teachers. Regarding supply types of assessment, they found that Beijing teachers used more of oral reading/dictation, oral discussion, short essays, taking notes, retelling stories, standardized writing and speaking tests, and translation than Hong Kong teachers, but used much less of student journal and student portfolio than Canada teachers and much less of long essay than Hong Kong teachers.

Regarding assessment purposes, Cheng, Rogers, and Hu (2004) categorized assessment purposes into three groups: *instructional purposes* such as "planning my instruction" and "grouping my students at the right level of instruction in my class", *student-centred purposes* such as "obtaining information on my students' progress" and "providing feedback to my students as they progress through the course", and *administrative purposes* including "providing information to the central administration" and "providing information to an outside funding agency" (ibid, p.367). It can be seen that these three types of purposes generally correspond to the three identities of assessment found in Rea-Dickins's (2001) paper, namely, the *instructional purposes* match the pedagogical identity, the *student-centred purposes* match the learning identity, and the *administrative purposes* match the bureaucratic identity of assessment. However, one purpose listed under *student-centred purposes* in Cheng et al.'s (2004) paper, "determine the final grades for my students", seems to be more related to *administrative purposes* than to *student-centred purposes*, since in most cases,

determining the final grades is summative in nature and the final grades are usually used for reporting purposes.

Regardless of their classification, their questionnaire survey results showed that teachers from Beijing focused more on assessment for instructional purposes while Hong Kong and Canadian teachers were more concerned with assessments for student-centred purposes, except that more Beijing teachers than teachers from Canada and Hong Kong would use assessment to make students work harder (Cheng et al., 2004).

In their interview study (Cheng et al., 2008), the researchers did not adopt the classification they used in their previous study (Cheng et al., 2004). Instead, they grouped all the purposes together and asked the interviewees to match the purposes with their selected assessment methods. Specifically, they found that Canadian teachers used selection types of assessment methods mainly for formative purposes such as to "obtain information on my students' progress", "provide feedback to my students as they progress through the course", and "diagnose strengths and weaknesses in my students". On the other hand, Beijing teachers used such assessment methods mainly to "obtain information on my students' progress" and to "prepare my students for standardized tests they will need to take in the future". Although these purposes are somewhat formative in nature, there was an obvious washback effect from external standardized testing. Regarding supply types of assessment methods, while all the three groups of teachers used their chosen assessment methods mainly for student-centred purposes, Beijing teachers were unique in two ways: they tended to use supply types of assessment methods to "motivate my students to learn", and use "standardized tests" to "prepare my students for standardized tests they will need to take in the future".

Regarding grading and reporting practices, Cheng and Wang (2007) found that Canadian teachers used analytical scoring[2] most, Hong Kong teachers used rubric

---

[2]Analytical scoring means teachers "give marks for different components of an essay or a spoken presentation" (Cheng & Wang, 2007, p.106).

scoring[3] most, and Beijing teachers used holistic scoring[4] most. Moreover, Canadian teachers preferred percentage scores, Hong Kong teachers preferred letter grades converted from percentages, and Beijing teachers usually gave a score out of 100. The authors attributed such differences to teacher beliefs about the importance of their assessment practices:

> Hong Kong teachers tended to focus on the issue of validity and reliability for assessment and thus emphasized the importance of its technical qualities. Canadian teachers were concerned with matching the purposes of assessment to individual needs and therefore valued the type of grading that was maximally useful for student learning. Teachers from Mainland China, although they did not demonstrate explicit knowledge about analytical and rubric scoring, seemed to prefer holistic scoring out of habit. (Cheng & Wang, 2007, p.101)

In addition, the authors argued that the "practicality of assessment" also played a role, since most Chinese classes were much larger than Canadian and Hong Kong teachers' classes, which might push Chinese teachers to seek the most convenient way to grade.

Regarding follow-up strategies after an assessment, Cheng and Wang (2007) found that most Canadian and Hong Kong teachers tended to provide individualized feedback, while Beijing teachers tended to provide feedback to the whole class partly due to the larger classes they were teaching. In addition, Canadian teachers tended to go over test answers or the scoring criteria with their students and review students' performance, while Hong Kong teachers tended to provide students with individual feedback sheets and ask students to do self-feedback and self-reflection. In contrast, Beijing teachers tended to read exemplar answers to the whole class.

Cheng and her colleagues' studies (Cheng et al., 2004; Cheng et al., 2008; Cheng & Wang, 2007) were valuable in presenting a comprehensive picture about the characteristics of three groups of tertiary-level ESL/EFL teachers' CA practices. Their studies also revealed a relationship between teachers' teaching contexts and their assessment practices. However, their results mainly depended on teachers' self-report data and they pointed out that future studies of assessment practices should include

---

[3]Rubric scoring means teachers "match essays or presentations to one of four performance descriptions that differ according to completeness and correctness" (Cheng & Wang, 2007, p.106).
[4]Holistic scoring means teachers "give one mark for overall impression" (Cheng & Wang, 2007, p.106).

"both teachers' perceptions and their actions in the classroom" (Cheng & Wang, 2007, p.103).

Unlike the above-mentioned survey studies, the following three studies investigated one specific aspect relating to teachers' CA practices, namely, teachers' judgment-making processes.

Davison (2004) conducted a comparative study to investigate how Australian and Hong Kong ESL teachers made assessment decisions when they were asked to mark six written argumentative essays produced by final-year secondary school students. Through analysis, the author proposed a framework which consists of five types of teacher-rater orientations: "technician," "interpreter of the law", "principled yet pragmatic," "arbiter of 'community values,'" and "assessor as God". The five orientations fall along a continuum with the "technician" orientation more criterion-bound and mechanical and the "assessor as God" orientation more intuitive with implicit constructs and interpretations (as seen in Figure 2.10).

Assessor as technician	Assessor as the interpreter of the law	Assessor as the principled yet pragmatic professional	Assessor as the arbiter of 'community' values	Assessor as God
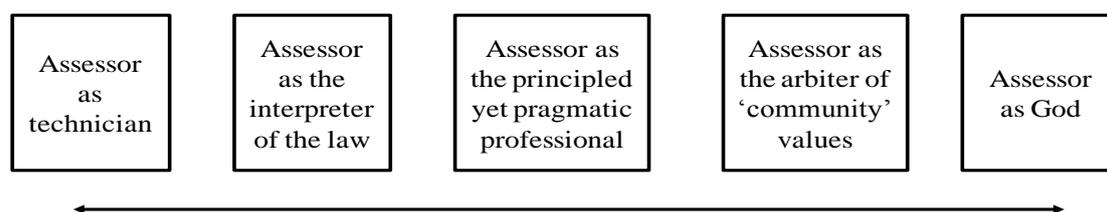
Figure 2.10 Davison's Classification of Teacher-Rater Orientations

Butler (2009) conducted a similar study at secondary school in South Korea and confirmed the usefulness of Davison's framework. By asking 26 elementary-school teachers and 23 secondary-school teachers to assess four 6th-grade students' performances during two group activities, Butler found differences among the two groups of teachers in terms of their views on assessment criteria, how to evaluate students' confidence and motivation, and how to measure students' potential ability to communicate competently in a foreign language. Analyzing such variability against Davison's (2004) framework, Butler found that the elementary teachers tended to be more oriented towards the "assessor as God" and "assessor as the arbiter of 'community values'" positions, while the secondary school teachers were more

oriented towards the "assessor as technician" and "assessor as the interpreter of the law" positions.

While the above two studies were situated in the primary- and secondary-school levels and examined teachers' judgment-making processes over one or two assessment activities, the following study was conducted at the university level and with teachers' CA practices over a wider range of assessment activities.

Yin's (2010) case study explored two English language teachers' cognition while engaging in classroom language assessment in a UK university language centre. Through classroom observation and repeated stimulated recall interviews, Yin identified a wide variety of interrelated cognitions underlying the two teachers' assessment practices, which were grouped under two categories according to when the cognitions mainly occurred. Strategic cognition occurred during the planning stage while the interactive cognitions were in operation mainly as teachers assessed students during class time. Strategic cognitions may be about "teaching approach and beliefs about language learning, classroom parameters, and course syllabus and summative assessment", which "substantially influenced teacher thinking in relation to assessment"; and interactive cognitions include "assessment principles, constructs applied interactively, stereotyping, projection, mental portraits of students, and assessment not directly related to language use", which were operative mainly as teachers assessed students during class time" (ibid, p.182). The study also found that a teacher's assessment practices are often influenced by many significant others, like teacher-trainers and managers of the language centre. Therefore, Yin argued that some teachers' "assessor as God" view of themselves (Davison, 2004) seems untenable because "their judgments of student language ability are potentially influenced by these other agents" (Yin, 2010, p. 192).

2.4.3 Relationships between CA and student learning

Several studies, all conducted in tertiary institutions, touched upon the relationship between certain aspects of teachers' CA or FA practices and student motivation or academic achievement. While some studies examined this kind of relationship by

looking at CA/FA practices that were already being used in a specific context (Colby-Kelly & Turner, 2007; Huang, 2010; Ross, 2005; Wang & Cheng, 2010), the others investigated the relationship by focusing on newly-designed CA/FA systems or practices (Birjandi & Tamjid, 2012; Cao, Zhang, & Zhou, 2004; Ishihara, 2009; Zhou & Qin, 2005). Whereas four studies specifically examined this kind of relationship by looking at students' linguistic improvement (Birjandi & Tamjid, 2012; Cao et al., 2004; Ishihara, 2009; Ross, 2005), the rest of the studies in this group mainly examined the impact of CA/FA practices on students' non-linguistic achievement such as their motivation and self-regulated learning. In what follows, those studies that focused on students' linguistic improvement are reviewed first.

Ross's (2005) study found that FA practices are more conducive to promoting improvement in students' listening proficiency than traditional summative assessments. In his longitudinal study, Ross collected data from eight cohorts of undergraduate students (n=2215) from a university in Japan regarding their English language learning and achievement over a period of eight years. The first four cohorts received traditional summative assessments (such as instructor-graded homework, quizzes, assignments, report writing projects, and end-of-term tests); and the latter four cohorts received more formative assessment practices (such as self-assessment, peer-assessment, on-going portfolios, and cooperative learning projects). Statistical analysis found that students receiving FA practices achieved significantly more proficiency growth than those receiving traditional summative assessments in the academic listening domain, but not so in the academic reading domain.

Two studies touched upon EFL learners' writing ability. Birjandi and Tamjid (2012), through a pre-test-post-test quasi-experimental study, found that self- and peer-assessment together with teacher assessment significantly improved learners' writing performance as compared with teacher assessment only. Working with 157 TEFL juniors in one university in Iran, the authors divided the students into five groups, each receiving one of the following five treatments for one semester:

Group 1: Journal Writing + Teacher Assessment (JW + TA) with 30 student

Group 2: Self-Assessment + Teacher Assessment (SA + TA) with 37 students

Group 3: Peer Assessment + Teacher Assessment (PA + TA) with 31 students

Group 4: Self-Assessment + Peer Assessment (SA + PA) with 29 students

Group 5: Teacher Assessment (TA) with 30 students (p.517)

Statistical analysis showed that students from Groups 2 and 3 had significantly higher mean scores than the students from the control group, Group 5, while the mean scores of the students from the other two groups, Groups 1 and 4, were not statistically different from the mean score of the control group students.

Similarly, another study (Cao et al., 2004) conducted in a writing course found a newly-designed assessment system helped improve EFL learners' writing ability. The assessment system consisted of three major components: 1) student portfolio, which collected information about student in-class learning (including essays written in class, notes taken in class, students' self-assessment and peer-assessment) and after-class learning (including weekly journals about their learning, goals set for improving their writing, and essays written after class); 2) student survey, which included students' demographic information, midterm and end-of-semester questionnaire survey and student reflection; 3) writing tests, which included a pre-test, a midterm test, and a post-test. This assessment system was implemented in an optional EFL writing course for ten weeks and 177 non-English-major undergraduates from one university in China participated in the study. Paired sample T-test between pre-test and post-test showed significant improvements in the three aspects in students writing: content, vocabulary use, and fluency. In addition, the study found much positive feedback about this assessment system from the students: 1) the majority of the students thought it could better reflect what students had learned from this course than the traditional final-test-only assessment method; 2) the portfolio method helped students see their own progress; and 3) some students demonstrated autonomy over their own writing process. However, students' feelings about self-assessment and peer assessment varied according to their language proficiency levels. One significant point about the study is that the FA results were also used for the final achievement scoring (i.e. for summative purposes), which probably had motivated the students. The generally positive feedback from the students showed that it was possible to integrate formative and summative assessments and make them both serve learning.

Ishihara's (2009) ethnographic case study reported a newly developed classroom-based assessment system that enhanced EFL learners' pragmatic competence. The assessment system consisted of ten sequenced assessment/instruction activities spread out through a whole semester: 1) initial reactions to language use in context, 2) production of written request discourses, 3) learners' data collection in authentic L1/L2 discourse, 4) learners' reflections on language use in context, 5) learners' analysis of context-language relationship, 6) pragmalinguistic development and assessment, 7) learners' self-revising, role-playing, and refining request discourses, 8) learners' self-evaluation of written request discourse, 9) teacher's assessment of written request discourses, and 10) teacher-learner collaborated assessment of intention-interpretation match. The author tried it out with her own EFL classes, 58 Japanese freshmen from three classes majoring in business management. This assessment system also served as tools for data collection. In addition, she also took field notes and asked students to do three questionnaires (initial student background survey, midterm reflective questionnaire, and course evaluations). Through analyzing learner language, she found that this assessment system developed learners' pragmalinguistic competence as well as their socio-pragmatic awareness of the consequences of their own pragmatic choice, but there were also individual differences due to students' own attitudes towards classroom instruction.

While the above-mentioned studies revealed that certain CA/FA practices can enhance students' listening proficiency, writing performance and pragmatic competence, the following studies obtained some findings on the relationship between CA/FA practices and students' motivation / self-regulated learning strategies.

Colby-Kelly and Turner (2007) investigated the relationship between AfL and student motivation by examining two speaking activities (debate and presentation projects) at a Canadian continuing education program specializing in pre-university English for Academic Purposes (EAP) classes in Canada. Through classroom observation and student interviews with 12 students from two advanced-level EAP classes in Quebec, they found that "teacher-student feedback with a motivational component appears to be useful and motivating to some learners and may benefit learning as a result, but may not be useful to those who don't take it seriously" (ibid, p.33).

In another study, Huang (2010) investigated the impacts of convergent vs. divergent assessment on college EFL students' motivation and self-regulated learning strategies in one teacher's listening and speaking classes in Taiwan. In this study, convergent assessment was represented by a more traditional listening test and an oral exam, and divergent assessment was represented by a group presentation from the listening class and a group presentation from the speaking class. A total of 105 college freshmen participated in the study and each student experienced one convergent assessment and one divergent assessment. After each student experienced each of the two assessments, they completed a questionnaire on their task-specific motivation and learning strategies. Analysis showed that students with lower self-efficacy demonstrated higher motivation and strategy use under divergent assessment while students with higher self-efficacy did so under convergent assessment. Furthermore, in the speaking class, student motivation and strategy use were higher for divergent assessment than for convergent assessment, but vice versa in the listening class. This study revealed that the impacts of CA practices on students might be mediated by students' own self-efficacy as well as specific language skill areas.

Zhou and Qin (2005) reported how their newly-designed assessment system helped enhance student motivation and learner autonomy. The assessment system they designed consisted of three components: 1) student portfolios (including self-evaluation of self-development, records of learning process, feedback from the teacher and classmates, and a collection of students' work such as their presentations, role plays, drama performance, etc.); 2) records of teacher observation (including classroom observation records, records of teacher-student conferencing, students' assignments, and messages left on the class webpage); and 3) evaluation of learning outcomes (including students' self-assessment, peer assessment, teacher assessment, and final test). This assessment system was implemented in two College English classes, altogether 78 students, at a university in China, one freshmen and one sophomore, all majoring in computer science. In addition to the information collected through the assessment system, such as student portfolios, teacher observation records, student self-assessment and peer assessment, and teacher assessment, the researchers also conducted a questionnaire survey of students' perceived impacts of the assessment system on their motivation, interest, learner autonomy, and corporation

ability at the end of the semester. The results showed that this assessment system not only stimulated learners' motivation and interest to learn, but also made it possible for the teacher to monitor and facilitate learners' learning processes as well as their autonomy. They also found evidence that the assessment system could cultivate learners' autonomy, ability to cooperate and communicative skills.

Another study in China explored the relationship between students' perceptions of classroom assessment environment and their goal orientations, a particular type of motivation. Based on Brookhart's concept and model of "classroom assessment environment" (Brookhart, 1997), Wang and Cheng (2010) operationalized the concept through questionnaire items on the features of teachers' assessment methods, such as grading harshness and task difficulty, and assessment feedback, such as personal relevance of feedback. Through a questionnaire survey administered to 503 first-year non-English-major undergraduates from one university in China, the authors found that when students felt the assessment environment was learning-oriented, they tended to have strong mastery goals; when they felt the assessment environment was test-oriented, they tended to have performance avoidance goals; and when they felt the assessment environment was praise-oriented, they tended to have performance approach goals[5].

2.4.4 Summary

The following features have emerged from the review of the empirical studies on CA/FA in L2 assessment field.

Studies conducted in EAL/ESL school contexts seem to be more concerned with the FA-SA relationship, while studies conducted in EFL tertiary-level contexts seem to be more concerned with portraying the existing CA practices or understanding the relationship between certain aspects of CA and student learning, linguistically or non-linguistically.

---

[5]Mastery goals emphasize working hard in order to develop one's skills, understand one's work, improve one's competence, or achieve a sense of mastery based on self-referenced standards. Performance approach goals emphasize working hard in order to outperform others, while performance avoidance goals emphasize working hard in order to avoid looking incompetent or stupid (cf., Ames, 1992; Maehr & Meyer, 1997; Middleton & Midgley, 1997)(Ames, 1992; Maehr& Meyer, 1997; Middleton &Midgley, 1997)

While many models on CA/FA have been proposed (cf. Section 2.3), few of the empirical studies have been informed by any of those models (except Wang & Cheng, 2010), which reflects a disconnection between theory and empirical research.

Teachers' CA practices have been understood mainly through questionnaire surveys or teacher interviews (e.g., Cheng et al., 2004; Cheng et al., 2008; Cheng & Wang, 2007; Wang & Cheng, 2010), or by examining teachers' practices during a limited number of selected assessment activities (e.g., Butler, 2009; Davison, 2004; Huang, 2010; Yin, 2010). There is still a need for an overall picture of how teachers' CA practices manifested themselves in classroom settings.

While researchers have examined the influence of CA/FA on learner achievement, both linguistically and non-linguistically, studies investigating CA practices in L2 speaking classrooms are still scarce, apart from those by Ishihara (2009) and Colby-Kelly and Turner (2007).

For those studies that examined the impact of newly-designed assessment systems on learning (Birjandi & Tamjid, 2012; Cao et al., 2004; Ishihara, 2009; Zhou & Qin, 2005), the researchers generally took an action-research approach, where they had total freedom to design and conduct assessment the way they thought best to serve student learning. Their assessment system did not seem to have been restrained in any way by any external factors such as assessment structures imposed from above. Therefore, more research is needed to reveal teachers' CA practices in their naturalistic settings.

2.5 Framework for the present study

Although the present study aimed to obtain an in-depth and contextualized understanding of CA practices in EFL speaking classrooms, which generally called for an *emic* perspective in data collection and data analysis, it would be unwise to conduct the research without considering what had been found in the field of CA. Therefore, to make the best use of the existing relevant literature and to remain open and sensitive to the specific characteristics of each case, the present study adopted a combination of *etic* and *emic* approaches.

From the *etic* perspective, this study was informed by the following theories and research. Regarding research design, Harlen's (2007) component-dimension model of CA (Figure 2.2), Black and Wiliam's framework of FA (cf. Section 2.3.3), and Brookhart's (1997, Brookhart et al., 2006) model of CA (Figures 2.4 & 2.5) suggested the important aspects and relationships to look at when researching CA. Broadly speaking, researching CA should involve understanding of the assessment methods, purposes, feedback, judgment-making processes, participants' roles in and perceptions of an assessment process. Moreover, Harlen's (2007) time-dimension model of CA (Figure 2.3) and Wiliam and Thompson's model of FA (Figure 2.6) indicated the time when CA might occur and thus suggested when to collect data concerning CA. These models especially drew the researcher's attention to everyday classroom teaching and to the teacher's feedback especially.

Furthermore, since the assessment construct is an important component of assessment as a system (Airasian & Russell, 2008; Bachman & Palmer, 2010; Banks, 2005; McMillan, 2007a), a comprehensive picture of a teacher's CA practices should not overlook the assessment constructs which are implicit in those CA practices. Considering the present study was situated in oral English classrooms, the author reviewed the existing models on L2 speaking ability and adopted Celce-Murcia's model (2008) of communicative competence as the starting point for analyzing assessment constructs of the identified CA practices in this study.

Writers on the assessment of L2 speaking ability (Fulcher, 2003; Hughes, 2011; Luoma, 2004) have broadly classified this ability into three levels. At the level of language production, speakers should be able to produce comprehensible speech with appropriate linking of words, stress, and intonation to convey the intended meaning. At the level of language choice, speakers should have a wide and appropriate range of vocabulary and grammatical resources to express ideas precisely. At the level of discourse, speakers need to have the ability to express ideas and opinions in coherent, connected speech. They should also have the ability to interact with the interlocutor by initiating and responding appropriately and at the required speed and rhythm. They should be able to use functional language and strategies to maintain or repair interaction. The three levels are clearly summarized in the ETS's (Educational Testing

Service, see www.ets.org) technical report on TOEFL Speaking Test (Xi, Higgins, Cechner, & Williamson, 2008) (see Figure 2.11).
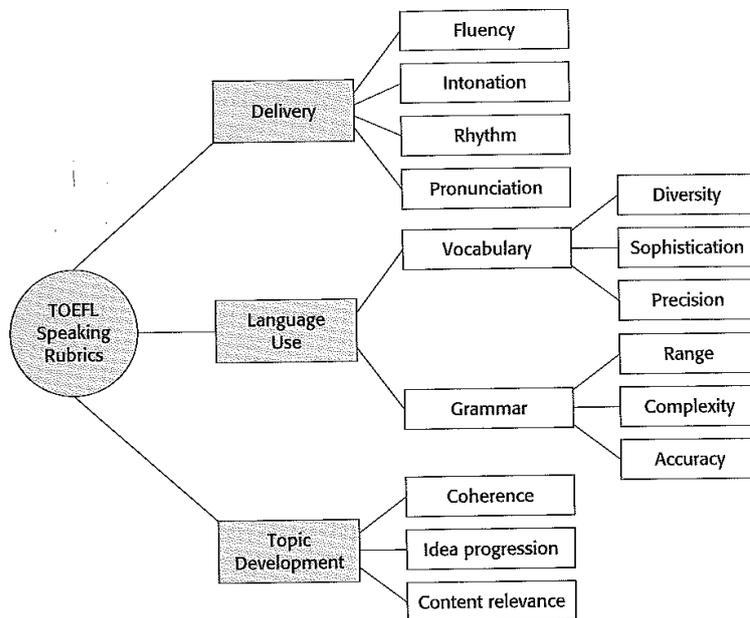
Figure 2.11 Categories Underlying the Speaking Construct (Source: Xi et al., 2008, p. 29)

With a slightly different approach, that is, with language teaching rather than language testing in mind, Celce-Murcia (2008) proposed a model to represent the construct of communicative competence (see Figure 2.12).
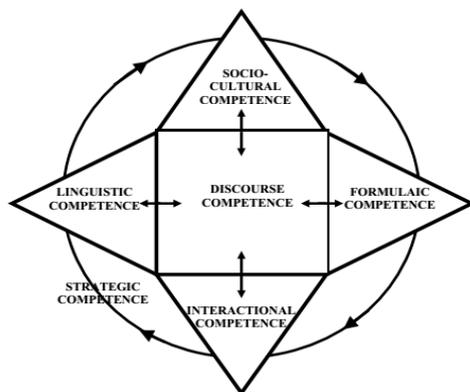
Figure 2.12 A Comprehensive Model of Communicative Competence (Source: Celce-Murcia, 2008, p.45)

Analysis of this model revealed its two characteristics. 1) It has the most number of components compared with previous models of communicative competence (cf. Celce-Murcia, 2008) and those aspects specified in Figure 2.11 are included in this model. Specifically, the "Delivery" and "Language use" components in Figure 2.11 are covered within "Linguistic competence" component in Figure 2.12; and the "Topic development" component in Figure 2.11 is similar to "discourse competence" component in Figure 2.12. Therefore, Celce-Murcia's model is more comprehensive. 2) It was designed for language teachers to be used in language classrooms. Since the present study was situated in actual language classrooms, this model should be informative for the present study. The following is a brief introduction of the components in this model. The key ideas in each component is underlined.

> **Sociocultural competence**: The speaker's pragmatic knowledge, i.e. how to express messages <u>appropriately</u> within the overall social and cultural context of communication
>
> **Discourse competence**: The selection, sequencing, and <u>arrangement</u> of words, structures, and utterances to achieve a <u>unified</u> spoken message
>
> **Linguistic competence**: The speaker's phonological, lexical, morphological, and syntactic knowledge
>
> **Formulaic competence**: The speaker's knowledge of those <u>fixed and prefabricated chunks of language</u> that speakers use heavily in everyday interactions.
>
> **Interactional competence**: This component has three sub-components.
>
>   *Actional competence*: The speaker's knowledge of <u>how to perform common speech acts</u> and speech act sets in the target language involving interactions such as information exchanges, interpersonal exchanges, expression of opinions and feelings, problems (complaining, blaming, regretting, apologizing, etc.), future scenarios (hopes, goals, promises, predictions, etc.)
>
>   *Conversational competence*: The speaker's knowledge of how to open and close conversations, how to establish and change topics, how to get, hold, and relinquish floor, how to interrupt, how to collaborate and backchannel, etc., which is inherent to the turn-taking system in conversation
>
>   *Non-verbal/paralinguistic competence*: The speaker's <u>body language, use of space, touching, and non-linguistic utterances such as Uh-oh</u>

Strategic competence: The speaker's learning strategies (cognitive, metacognitive, and memory-related) and <u>communication strategies</u> (achievement, stalling, self-monitoring, interacting, and social) *(emphasis added)* (Celce-Murcia, 2008, pp.46-50)

For the present study, the author adopted this comprehensive model as a starting framework to analyze the assessment constructs of the teachers' CA practices.

While the above-mentioned models and frameworks provided a general approach and some guidelines about how to go about this study, the ultimate aim of this research required the researcher to gain an insider's view of CA, that is, from the participant teachers' and students' view, rather than a single *etic* perspective of largely the researcher's point of view. Therefore, the researcher employed multiple research methods and kept an open mind during the study in order to reveal the participants' understanding and experiences of CA, the *emic* perspective, which will be described in detail in next chapter.

2.6 Chapter summary

Stemming from formative evaluation (Scriven, 1967) and CRM (Glaser, 1963), CA research began to attract increasing attention in both the general education field and the language assessment field in particular from late 1990s, especially after the publication of Black and Wiliam's seminal paper (1998a) which pointed out the importance of FA in promoting effective learning.

Over the past decade or so, with the importance of CA for teaching and learning being increasingly recognized, a considerable amount of literature on CA, including theoretical frameworks and empirical studies, has been published in the mainstream education field. However, in contrast, CA research in the L2 assessment field has fallen behind. As Davison and Cummins (2007b, p. 415) pointed out, "in many ways ELT has lagged behind the rest of the educational field in exploring new theories and methods of assessment and evaluation."

While CA has been found to be a multi-faceted, multi-dimensional, and multi-functional tool in classroom teaching and learning and many models have been

proposed, existing empirical studies have been generally uninformed by those models and only examined certain aspects of CA, such as assessment purposes (e.g., Rea-Dickins, 2001, 2007b), assessment methods (e.g., Davison, 2004), and its relationship with student learning (e.g., Ross, 2005). Few studies have examined CA as a whole system, as suggested by Harlen (2007).

In Chinese EFL contexts, while some researchers have tried out some ideas borrowed from western countries (Cao et al., 2004; Zhou & Qin, 2005) or have investigated teachers' CA practices through questionnaires and interviews (Cheng & Curtis, 2010a; Cheng, Rogers, et al., 2004; Cheng et al., 2008; Wang & Cheng, 2010), there is still a lack of descriptive information about teachers' actual CA practices in their own classrooms. Moreover, few studies have examined the relationship between FA and SA in a Chinese EFL context.

In light of the above literature review, this thesis, informed by the existing models/frameworks, set out to gain a holistic and contextualized understanding of teachers' CA practices in tertiary-level EFL speaking courses in China.

Chapter 3. Methodology

Ellis (2012) points out two principal research paradigms for classroom-based L2 research: the normative paradigm that "seeks to test hypotheses drawn from an explicit theory of L2 teaching or learning and typically involves some form of experiment," and the interpretive paradigm that "seeks to describe and understand some aspect of teaching by identifying key variables and examining how they interrelate" (p.x). The present study obviously fits into the interpretive paradigm because it sought to describe the participant teachers' CA practices in depth, holistically, and in context, so as to reveal the typical patterns of these practices in their EFL speaking classrooms.

To achieve such a purpose, this study adopted a multiple-case study approach. This chapter, after the presentation of the research questions, will discuss the rationale for the case-study approach. What follows is a description of the identification of the cases, the data-collection instruments and procedures, and the process of data analysis.

3.1 Research questions

This study aimed to uncover and crystallize EFL teachers' CA practices in naturalistic classroom settings. As discussed in chapter 2, CA, as an umbrella term, is not only about grading students but also about enhancing teacher teaching and student learning. In other words, CA includes both SA and FA. While SA practices should be salient in this EFL context since the educational culture in China is still examination-dominant (Carless, 2011; Cheng & Curtis, 2010b), this study also made an effort to capture the teachers' FA practices, especially those remained oblivious to the teachers but fit the definition of CA. Therefore, this study addressed the following research questions.

1)   How did the teachers assess their students in their oral English courses?

   1a) What were the recognized CA practices, including both SA and FA, in these oral English courses? What was each practice like?

   1b) Were there any unrecognized CA practices in the teachers' oral English courses? If yes, what were they like?

2) Why did the teachers assess their students the way they did?

3) What were the students' perceived impacts of their teachers' CA practices?

3.2 Rationale for the case-study approach

In essence, this study was an exploratory and descriptive study, aiming to uncover and crystallize EFL teachers' CA practices in their own oral English classrooms in China. Because the literature suggests that CA is a multi-faceted multidimensional phenomenon, performing multiple functions in localized classroom contexts, a case study approach is the best choice to capture such a complex phenomenon, thanks to its advantages such as allowing contextualized, in-depth and holistic understanding of a complex issue and allowing for mixed methods and multiple research paradigms (Bassey, 1999; Dornyei, 2007; Mackey & Gass, 2005; Merriam, 1988; Patton, 2002; Yin, 2009). As Nunan and Bailey (2010, p. 158) points out, a case study is a "detailed, often longitudinal, investigation of a single individual or entity", and "as a type of naturalistic inquiry," it does not "involve any sort of treatment" but allows the researcher to "learn what is happening." These features just serve the purpose of the present study well.

The decision to carry out a multiple-case rather than a single-case study was based on Yin's (2009) suggestion as well as the real situation in China. Yin (2009) said "single-case designs are vulnerable … because you will have put 'all your eggs in one basket.' More important, the analytic benefits from having two (or more) cases may be substantial" (p.61). In China, although oral English is widely offered to university students, there is a wide range of variations from university to university and from department to department concerning textbooks used, course requirements, course syllabus, students' English levels, teachers' teaching experience, etc.. Therefore, it was difficult to justify any one case as a representative case for this particular context. Consequently, a multiple-case approach could allow the researcher to select a few cases in contrasting situations, which could enhance the generalizability of the study findings (Dornyei, 2007).

3.3 Identifying the cases

In case studies, defining the *case* is of paramount importance (Simons, 2009). Although traditionally a case is defined as an entity, such as a person, a classroom, a policy, a process, etc., researchers have pointed out decisions about case selection also concern specific aspects of the selected entity to be investigated, since it is impossible to study every aspect of a selected entity (Duff, 2008; Yin, 2009). Since a case is a "bounded instance" (Nunan & Bailey, 2010, p. 161), it is important to set boundaries so that the study remain focused and how the study is related to readers can be made explicit (Merriam, 1998; Yin, 2009). Following Yin's (2009) suggestion that the defining of a *case* should be directed by the research purpose and research questions, the present study regarded a university EFL teacher's CA practices in his/her oral English course in China as the unit of analysis in this study. More specifically, this study examined: EFL teachers, not ESL or EAL teachers; tertiary-level teachers, not primary school or middle school teachers; oral English teachers, not teachers teaching other subjects such as reading or writing; the teacher's CA practices, not his/her teaching strategies, although some of his/her CA practices might also function as teaching strategies such as incidental assessment opportunities (Hill, 2012).

At tertiary level in China, there are various types of universities such as comprehensive universities, foreign language universities, universities of science and technology, universities with specialties such as law, forestry, and medicine. In spite of such variations, almost every university has both English-major programs, where students have English as their major and spend most of their time learning and perfecting their English language skills, and non-English-major programs, where students have their own majors and they spend much less time learning English. In addition, while most universities still employ Chinese teachers to teach oral English, more and more native English speakers are being hired to teach oral English, especially to English-major students. Moreover, with society valuing a person's oral communication skills more and more in China, not only English-major students but also non-English-major students attach great importance to their oral English courses.

Bearing such variations in mind, the researcher made an attempt to carry out "maximum variation sampling" (Miles & Huberman, 1994, p. 28). She approached 16 teachers from ten different universities in Beijing. In the end, given the strong commitment this study required on the part of the participants, three teachers from two universities were recruited. As explained above, the intention was to recruit teachers who varied according to their personal characteristics, their educational background and the kind of courses they were teaching to achieve maximum variation sampling. As the profiles in Table 3.1 show, this variation was achieved to a large extent.

Table 3.1 Teacher Participants' Background Information

Pseudonym	Andrew	Mary	Linda
Nationality	Chinese	American	Chinese
Gender	Male	Female	Female
Age group	In his early 30's	In her late 50's	In her 40's
University type	Foreign language university	Comprehensive university	Foreign language university
Student type	Non-English majors	English majors	English majors
Oral English course	Year 1 Semester 1 oral conversation	Year 1 Semester 1 oral discussion	Year 1 Semester 2 public speaking
Teaching experience	7 years	11 years	22 years
Educational background	B.A. in English education (4 years at a Chinese university) M.A. in American literature (2 years at a Chinese university)	B.A. in Spanish language at an American university M.A. in TESOL (1 year at an American university)	B.A. in English language and literature (4 years at a Chinese university) M.A. in TEFL (18 months at an American university) Ph.D. in applied linguistics (4 years at a New Zealand university)

3.4 Data collection

To address the research questions, the present study adopted a variety of data-collection methods, mainly qualitative in nature, conducted throughout the whole

semester of the selected courses. Miles and Huberman (1994) pointed out that qualitative research has three strengths: a strong handle on what 'real life' is like, a strong possibility for understanding latent, underlying, or non-obvious issues, and strong potential for revealing complexity (p.10). Such strengths serve the purposes of the present study well, that is, to gain an in-depth and comprehensive understanding of a complex and contextualized phenomenon.

A small-scale one-case pilot study was first conducted in January 2010, aiming to check and refine the data-collection instruments and rehearse the procedures for carrying out such fieldwork studies. The pilot study was conducted in an advanced-level ESL speaking course in a New Zealand university from Jan. 5, 2010 to Jan. 19, 2010. This was a summer course, with four sessions each week (100 minutes for each session). Every four sessions of this 24-session intensive course were devoted to one unit. After each unit, two sessions were scheduled for an end-of-unit formal assessment. This pilot study focused on the first unit of the course, together with its end-of-unit formal assessment, which altogether lasted for one week and a half. This class was very small, with only six students, all immigrants from Asian and Pacific countries. The teacher, a native New Zealander, and the six students participated in the pilot study. Based on this pilot study, revisions were made concerning the instructions for student journals, the interview guides, and the way of conducting stimulated recall, which, to avoid misunderstanding, was renamed as stimulated retrospective interviews because they were not conducted the way a typical stimulated retrospective interview is usually conducted (Gass & Mackey, 2000). These revised and improved versions of the instruments were used in the main study, which will be described below.

In what follows, I will first describe each of the data-collection instruments used in the main study, and then describe the data-collection procedures at each site.

3.4.1 Data-collection instruments

Richards and Morse (2007), when discussing the design of a qualitative study which will be as valid as possible, pointed out that the researcher should try to achieve a high level of alignment between the research questions, the data collected, and the methods used to collect the data. Therefore, guided by the research questions and existing

literature and taking into consideration the actual settings, this study collected four types of data through four data-collection instruments (see Figure 3.1), as an effort to maximize the validity of the present study.
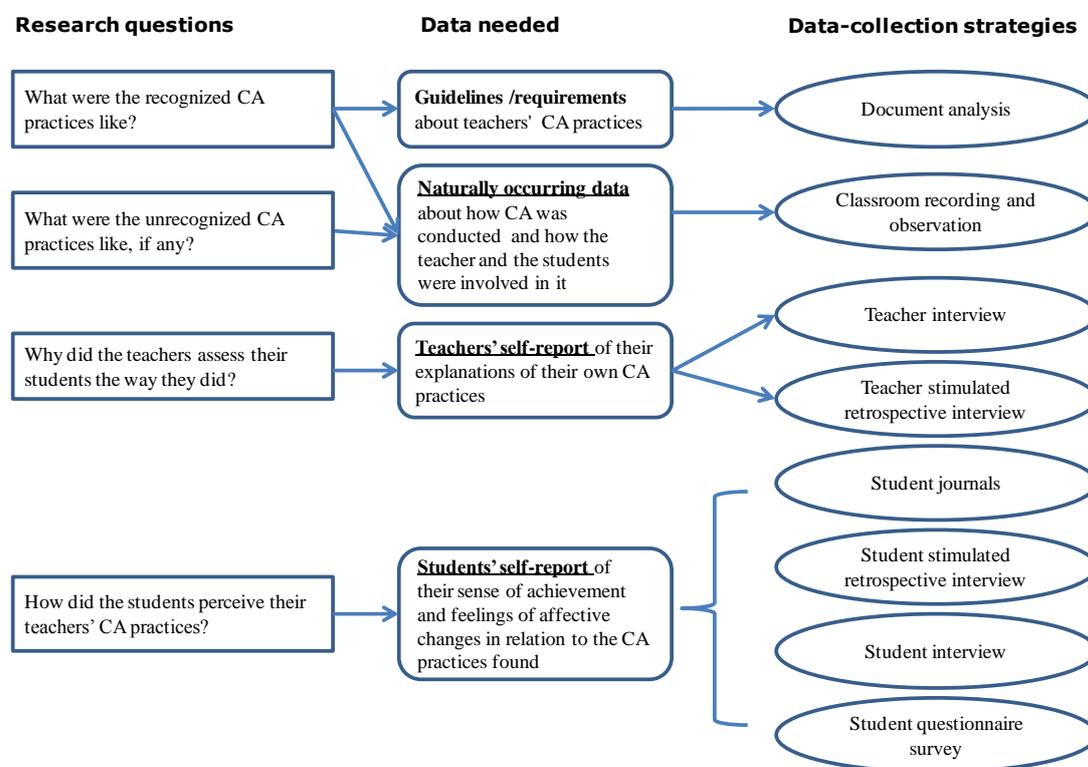
Figure 3.1 The Match between Research Questions, Data Needed and Data-collection Strategies

In what follows, the data collection instruments will be described in turn.

Documents

Documents were a valuable source of data "not only because of what can be learned directly from them but also as stimulus for paths of inquiry that can be pursued only through direct observation and interviewing" (Patton, 2002, p. 294). This research method was considered valuable for the present study due to the nature of this study: an exploratory and descriptive study of EFL teachers' CA practices in context. As CA practices are always contextualized, the researcher analyzed the following publicly accessible official documents including the National Curriculums for English Majors

and Non-English Majors and website information about each teacher's school and university, paying special attention to the university/school guidelines and/or principles regarding assessment of student learning. The analysis of this group of documents helped situate the present study in a larger context, especially what was mandated concerning CA, which served as a basis for the practices that actually happened in classroom context.

In addition to analyzing the publicly accessible documents, the researcher also analyzed the teachers' textbooks, course descriptions, teacher group meeting minutes, test papers, marking schemes, lesson plans, if available. The analysis of this group of documents not only provided a starting point for the identification of the teachers' CA practices, but also informed other methods of data collection such as the fieldwork observations, the interview questions posed to the participants, and student journal topics. They were also used together with data obtained from observations and interviews to reveal the connection between stated and actual CA practices.

Classroom recording and observation

To capture the instruction-embedded FA practices, many researchers have adopted the method of classroom discourse analysis which helped them analyze classroom recording data and generate patterns from the analysis (Cowie & Bell, 1999; Leung & Mohan, 2004; Rea-Dickins, 2001, 2006; Torrance & Pryor, 1998). As the present study took a very broad view of CA, in order to reveal all the existing CA practices, including those embedded in daily teaching and learning activities, the researcher recorded some of the participant teachers' lessons, which served as the working data for discourse analysis in order to reveal the hidden CA practices. To strike a balance between the maximum amount of data collected and the minimum interference on a teacher's normal teaching, the researcher recorded four weeks' lessons from each teacher, including daily teaching activities as well as part of the midterm test in Mary's case and some marked assignments in Linda's and Andrew's cases.

While video-recording is more powerful than audio-recording in uncovering the subtle reality of classroom life, it tends to cause more anxiety on the participants' part and consequently may distort the natural data (Zuengler, Ford, & Fassnacht, 1998).

Therefore, to make the data-gathering less intrusive, audio-recording instead of video-recording was adopted. During the observed lessons, the researcher usually came earlier and put a recorder inside the teacher's pocket or on the teacher's desk or near where the teacher would usually sit, to ensure that the teachers' voices were clearly recorded. The researcher did not put another recorder among the students so as to minimize any potential anxiety that might be caused by the recorder. As the recorder is guaranteed to capture sounds within 40 meters, it successfully captured the sounds when students talked individually in class, but when students talked in small groups, sometimes it was too noisy to distinguish who was talking what.

To complement the audio-recordings, the researcher also conducted classroom observation. It has been pointed out that the way to conduct classroom observation can vary along two dimensions. Firstly, the researcher's role in the classroom may vary from "participant" to "non-participant" (Dornyei, 2007, p. 177). Because the purpose of the present study was to obtain naturally occurring data, the "non-participant" stance was adopted. The researcher tried her best to minimize the influence of her presence on the observed classes by observing classes from a position located towards the back of the room, close to students but not so close to attract undue attention.

In addition, the observation data also vary from "structured" to "unstructured" (ibid, pp. 177-178). Considering that FA practices may be highly integrated within classroom teaching activities and may vary in form, time and size (cf. chapter 2), no *a priori* classroom observation scheme, such as the Communicative Orientation of Language Teaching (COLT) (Spada & Fröhlich, 1995), was used. Instead, a loosely-structured observation instrument was used, which consisted of four columns: time (starting time and ending time of a classroom activity), teacher's words and behaviours, students' words and behaviours, and the observer's comments and reflections (Appendix 1). The time column helped trace the classroom activities on my recording. The two middle columns were used to remind the researcher of what an activity or an interaction episode was about afterwards. The last column was the researcher's real-time judgment of whether an activity or an interaction episode was potentially a CA practice. If an activity was noted as a potential CA practice according

to the field notes, the researcher generally paid more attention to this activity during the follow-up stimulated retrospective interviews.

Teacher interviews

To find out teachers' explanations of their own assessment practices, teachers' self-report data about their beliefs on teaching, learning, and assessment, as well as their explanations of their own CA practices should be sought. To obtain such type of data, the interview has been a widely used method.

The interview is considered a highly interactive means of gathering data and is valuable because not everything can be observed (Fontana & Frey, 2003; Patton, 2002). It is not possible to directly observe feelings, thoughts, and intentions. Nor is it possible to observe "how people have organized the world and the meanings they attach to what goes on in the world" (Patton, 2002, p. 341). Therefore, interview data can supplement and triangulate with observation data. Out of this concern, interviews with the teachers were conducted in this study.

In this study, three interviews were conducted with each participant teacher. First, a baseline interview was conducted at the beginning of the semester to find out each teacher's educational background, teaching contexts, current teaching practices, and beliefs on teaching, learning, and assessment. At the end of this interview, the arrangements for classroom observations were also made (Appendix 2).

Second, at the end of the semester, each teacher was interviewed again to find out how they would summarize and reflect on his/her overall teaching experiences, with special attention given to their CA practices (Appendix 3).

Both the baseline interview and the end-of-semester interview followed a semi-structured interview guide for the following reasons. First, by listing the key questions and issues to be explored in an interview, the guide ensured that the most important issues were covered during the limited time available in an interview situation. Second, because this was a multiple case study, the guide might "ensure that the same topic lines of inquiry are pursued with each person interviewed" (Patton,

2002, p. 343). Third, although the key issues had been listed in the guide, the researcher still had the freedom to explore, probe, and ask questions that could illuminate this complex phenomenon: teachers' CA practices.

In addition to the above two interviews, each teacher had a stimulated retrospective interview right after one observed lesson. The purpose was to capture the teacher's thought processes involved in carrying out a task or activity where assessment was involved.

Generally, to obtain this kind of data, stimulated recall (SR) is a commonly used method (Gass & Mackey, 2000). However, during the pilot study the researcher noticed that almost all the activities in the oral English course might involve assessment and it was very time-consuming to conduct each SR interview by replaying the recording for each of the 100-minute sessions. Therefore, during the interview, the researcher just summarized all those activities or episodes according to the field notes and used the summaries as cues rather than replaying the recordings, which was somewhat different from the suggested practices for doing SR interviews (Gass & Mackey, 2000; Lyle, 2003). Therefore, it should be more appropriate to call this method a stimulated retrospective interview. Appendix 4 describes how the researcher prepared each teacher for this kind of interview and presents the generic questions asked during such stimulated retrospective interviews with each teacher.

Student questionnaires

To answer the third research question, students' self-report data were needed about their sense of achievement and their feelings about how the CA practices adopted in their courses affected their motivation, self-efficacy, effort, and learning. Considering the fairly large number of students in each class, three ways were adopted to collect data from the students: student questionnaires, student journals, and student interviews. This part will describe student questionnaires only. The other two methods will be described afterwards.

At each site, all the students from the selected teacher's oral English class were asked to complete two questionnaires, one at the beginning of the semester and one near the

end. The two questionnaires were used in this study for the following two reasons. First, because the researcher was only able to follow a limited number of students extensively in each setting, the convenience of a questionnaire survey (Bryman, 2001; Dornyei, 2003; Neuman, 2003) enabled the researcher to obtain data from more students, and thus get a more comprehensive picture. Second, the questionnaire data could complement and triangulate with those qualitative data obtained through classroom observations, student journals, and interviews, and thus enhance the internal validity of the study.

Aiming to get baseline information about the students, the beginning questionnaire (Appendix 5) consisted of two main parts: students' demographic information and their perceptions of their oral English learning including their self-evaluation of their oral English ability (items 1-7), their goal orientations in learning oral English in the selected course (items 8-16)[6], their anxiety in speaking English (items 17-19) and their effort in improving their oral English (items 20-23). The second part of the questionnaire was derived from empirical studies (Brookhart & DeVoge, 1999; Brookhart & Durkin, 2003; Brookhart et al., 2006; Wang & Cheng, 2010), which suggest that CA practices may influence students' perceived self-efficacy, motivation (specifically goal orientations), anxiety, effort, and overall achievement. Because students' perceived self-efficacy is usually associated with specific assessment tasks (Brookhart & Devoge, 1999) and the questionnaire for the present study was not directed towards any specific assessment practices, items on self-efficacy were not included in the questionnaire. For the second part, students were asked to indicate the degree to which they felt each statement was true on a five-point Likert scale (1 = strongly disagree, …, 5 = strongly agree). This part appeared both in the beginning questionnaire and in the end-of-semester questionnaire in order to see if students reported any changes in these aspects over one semester.

Near the end of their oral English course, all the students were asked to do a questionnaire survey again (Appendix 6). This end-of-semester questionnaire consisted of three parts. In addition to the two parts contained in the beginning

---

[6] Mastery goals: Items 8, 12, 15; Performance approach goals: Items 9, 11, 14; Performance avoidance goals: Items 10, 13, 16

questionnaire, this questionnaire had an additional part to find out students' perceptions of some typical CA practices, their frequencies and their usefulness to their learning. The typical CA practices were designed according to Wiliam and Thompson's (2008) model of FA (Figure 2.6). Specifically, this part tried to find out if teachers clarified and shared with students the learning intentions and criteria for success (items 1-5 and 20-22), how teachers elicited evidence of learning (items 6-8, 18, 19), how teachers provided feedback to students (items 9-15), and whether teachers involved students in CA (items 16, 17). Students were also given the space to add other CA practices that were not included in the questionnaire. The inclusion of this part III was intended to not only elicit students' general impressions of the CA practices in the selected courses, but also to see if there might be some relationship between their perceptions of the CA practices and their motivation, anxiety, effort, or overall self-evaluation. Such findings could also be used to triangulate with findings obtained through qualitative methods such as classroom observations, student interviews, and student journals. At the very end of the questionnaire, students were also invited to express their opinions and feelings about their oral English course.

English versions of the two questionnaires were tried out in the pilot study. However, because of the small number of student participants (only 6 students) and the short time interval between the two questionnaires (only one week and a half), only the wording of some items was improved based on the respondents' feedback.

Since the present study was conducted in China, to ensure that students understood the questionnaires accurately, the researcher translated all the questionnaires into written Chinese and then asked a Chinese colleague who is proficient in English to translate them back into English. The translated versions and the original versions were compared and all the discrepancies were discussed and revised to make sure the items in Chinese were accurate translations of the items in English.

Before being used in my main study, the Chinese version of the end-of-semester questionnaire was tried out with 48 first-year English-major freshmen in one university in China. Because of the time constraint, the researcher was unable to try out both questionnaires with the right time interval in between, so only the

end-of-semester questionnaire was tried out since it incorporated the initial questionnaire, aiming to check if the wording of the items was clear to the potential students. Based on the respondents' feedback, the researcher made revisions accordingly. Appendixes 5 and 6 contain the two revised questionnaires used in the main study.

However, the analysis of the questionnaire data from Andrew's and Mary's classes showed that on the one hand, the data were not suitable for inferential analysis due to the small number of students from each class, and on the other hand, students from the same class gave very different responses regarding their teacher's classroom practices, indicating that the questionnaire failed to provide an accurate picture of a teacher's CA practices (cf. Table 4.13). Therefore, the end-of-semester questionnaire was further modified (Appendix 7), aiming to get more straightforward answers from the students about their understanding of what constituted CA practices in their class (Part II), their perceptions of improvement, if any, regarding their overall self-evaluation, goal orientations, anxiety, and effort (Part III), and if they perceived any relationship between CA practices and their improvement (Part IV). Most of the CA practices listed in Parts III and IV were the same as those in the previous version (Appendix 6) but the revised items were specifically related to Linda's classroom activities. This revised questionnaire was used in Linda's class.

Student journals

To go deeper into students' perceptions of how they had been engaged in daily CA practices and their perceived impacts of such practices on themselves, volunteer students from each selected class were invited to write journals after each observed lesson.

Journal writing as a data-collection method usually refers to "solicited diaries", that is, accounts produced specifically at the researcher's request by an informant (Bell, 1999; Dornyei, 2007; Mackey & Gass, 2005). This method allows the researcher "an unobtrusive way of tapping into areas of people's lives that may otherwise be inaccessible" and allows the researcher to "study time-related evolution or fluctuation within individuals by collecting data on many occasions from the same individuals"

(Dornyei, 2007, p.157). Since the present study lasted for one semester, inviting students to write journals was thought to be an appropriate and effective way to accomplish this research purpose.

In the pilot study, the researcher tried out a journal template that consisted of three prompts: the students' experiences and feelings of the classroom activities in an observed lesson, what the students had learned about how to improve their speaking skills, and what had brought about such learning. After talking with the student participants and analyzing the data collected in the pilot study, the researcher realized that students could not recall a lesson clearly and their comments were very general and sometimes vague. In order to elicit more specific responses from the students, for each journal the researcher gave students a list of questions concerning particular activities based on the classroom observation and required students to give more specific responses concerning their feelings. An example can be found in appendix 8. It should be noted that the journal topics were written in English with Chinese translations, and students had the freedom to respond either in English or in Chinese.

Student interviews

Based on a similar reason to that for using student journals, some volunteer students also had a stimulated retrospective interview after one observed lesson. The way these interviews were conducted was similar to the stimulated retrospective interviews with the teachers. Appendix 9 describes how the researcher prepared each student for this kind of interview and presents the generic questions asked to each student.

In addition, the volunteer students were also interviewed at the end of the semester, to find out their overall impressions about this course, their perceptions of improvement and the contributing reasons, and their understanding and their feelings of the CA practices in this course (Appendix 10).

Summary

In this research, data were collected from various sources: from classroom observation and transcribed classroom discourse; from relevant documents; from interviews with

the case teachers and the volunteer students; from student journals and students'
responses to the questionnaires. The multiplicity of methods and sources of data
collection allowed the researcher to have access to the *emic* perspective as well as to
verify the researcher's interpretations (McKay, 2006). The next part will describe how
these methods were employed at each site.

3.4.2 Data-collection procedures

To capture the complexity and dynamics of CA practices in the selected oral English
courses without affecting the normal teaching too much, the following data-collection
procedures were carried out at each of the three sites (Figure 3.2).

At the beginning of the semester

| To obtain baseline information about the participants and the course | ✓ Conduct teacher baseline interview
✓ Conduct student baseline questionnaire
✓ Obtain relevant documents about the course |

During the semester

| • To describe and understand teachers' daily assessment practices;
• To find out how teachers' daily assessment practices are related to final summative grading;
• To understand students' perceived impacts of their teachers' daily assessment practices | ✓ Conduct classroom observation with recording for four weeks' lessons
✓ Invite volunteer students to write journals after the observed lessons
✓ Conduct teacher stimulated retrospective interview after one observed lesson
✓ Conduct volunteer student stimulated retrospective interview after one observed lesson
✓ Obtain additional relevant documents |

At the end of the semester

| • To describe and understand teachers' final assessment practices;
• To find out teachers' beliefs about teaching, learning, and assessment;
• To find out students' perceived impacts of the assessment practices in the course | ✓ Observe the final test
✓ Conduct student end-of-semester questionnaire
✓ Conduct teacher end-of-semester interview
✓ Conduct volunteer student end-of-semester interview
✓ Obtain additional relevant documents |

Figure 3.2 Data Collection Procedures at Each Site

At the beginning of the semester, a teacher interview and a student
beginning-of-semester questionnaire survey were conducted to obtain baseline

71

information about the participants. The teacher baseline interview was carried out at each teacher's own office. At the end of the baseline interview, an arrangement was made concerning when to do the classroom observations and the stimulated retrospective interview. At the end of the first week's lesson, with the permission of the teacher, the researcher went to the students' classroom, explained the study to the students and obtained their written consent before asking them to do the baseline questionnaire. Volunteer students who would participate in the journal writing and student interviews were also recruited then.

During the semester, four weeks' lessons were observed and audio-recorded as arranged. Immediately after each observed lesson, journal topics based on the content of the observed lesson were sent to the volunteer students through email. Students were asked to email their journals to the researcher within two days. Within one or two days of the observed lesson, stimulated retrospective interviews were conducted with one or two volunteer students in their own classrooms. After one of the four observed lessons, a stimulated retrospective interview was conducted with the teacher in his/her own office. Throughout the whole semester, every volunteer student and the teacher had one stimulated retrospective interview after one observed lesson.

It can be seen that the four observations together with all the interviews and student journals were like four snapshots of the teacher's daily classroom teaching practices. The reason behind this decision was that, although such snapshots could not capture the teacher's every assessment practice, they should provide sufficient amount of data to document the general pattern that the teacher usually followed. In addition, leaving a few weeks between observations allowed the researcher some time to reflect and adjust the data-collection strategies accordingly, to do preliminary analysis along the way, and also to collect data from the other cases. Therefore, to grasp the general pattern while exerting minimum disturbance to the normal teaching, four lessons of each teacher were observed and recorded.

Originally, the four classroom observations were planned to be spread out throughout the semester of a course. However, some actual difficulties during the data-collection stage of the study caused some changes. In Mary's case, since she did not have a

written syllabus, she preferred that two consecutive lessons were observed to reflect how her lessons were connected. In Andrew's case, since it took two weeks' lessons to cover one unit in his textbook, he also preferred that classroom observations should cover complete units. In addition, Mary and Andrew had their oral English classes on the same morning each week, but the two universities were far from each other, which made it impossible to observe both teachers' classes on the same day. Therefore, classroom observations for Andrew's case and for Mary's case were not evenly spread out, and Andrew's final test was not observed.

At the end of the last lesson, the student questionnaire survey was conducted in the students' own classroom. The final test was also observed (except for Andrew's case). After the final test, end-of-semester interviews were conducted with the teacher and the volunteer students. Table 3.2 on the next page shows the time when the data collection instruments were used at the three sites.

It should be noted that the student interviews were conducted in Chinese for most students, but two students from Linda's class and one student from Mary's class preferred to use English during their interviews. The teacher interviews were conducted mostly in English, but with Andrew and Linda, sometimes the Chinese language was used.

Approval for both the pilot study and the main study was granted by the Human Participants Ethics Committee at the University of Auckland in October 2009 and in March 2010. Details concerning the ethical aspects of this thesis are held at the University of Auckland under the approval numbers: 2009/436 and 2010/065.

Table 3.2 Data Collection at Each Site

Teacher	Teacher baseline interview time	Student baseline questionnaire time	Classroom observation time	Teacher retrospective interview time	Number of student retrospective interviews after each observation	Number of student journals after each observation	Time for student end-of-semester questionnaire	Time for teacher end-of-semester interview	Time for and number of student end-of-semester interview
Mary	Sept. 14, 2010	Sept. 14, 2010	Oct. 26, 2010		1	3	Dec. 21-28, 2010 (Students did it after class.)	Dec. 28, 2010	Dec. 28, 2010 to Jan. 7, 2011 4 students
			Nov. 2, 2010	Nov. 2, 2010	1	4			
			Nov. 23, 2010		1	3			
			Dec. 21, 2010		1	1			
			Dec. 28, 2010 (final test)						
Andrew	Oct. 14, 2010	Oct. 19, 2010	Nov. 9, 2010		1	5	Dec. 14, 2010 (Students did it at the end of that day's class.)	Dec. 21, 2010	Dec. 22-26, 2010 5 students
			Nov. 16, 2010	Nov. 16, 2010	1	5			
			Dec. 7, 2010		1	3			
			Dec. 14, 2010		2	0			
			Dec. 21, 2010 (final test)						
Linda	Feb. 25, 2011	Feb. 28, 2011	Feb. 28 & March 3, 2011		1	12	May 26-31, 2011 (Students did it after class.)	July 1, 2011	June 16-25, 2011 7 students
			March 28 & 31, 2011	April 1, 2011	2	8			
			April 25 & 28, 2011		1	3			
			May 23 & 26, 2011		2	5			
			June 16, 2011 (final test)						

3.5 Data analysis

To facilitate analysis within and between the three cases, all the recorded data were transcribed verbatim (transcription notation can be found in Appendix 11). Then all the data from one site were grouped into one big folder which consisted of seven sub-folders: folders 1-4 contained data about each of the four observed lessons (each including the observation field notes; transcript of the recorded lesson, teacher stimulated retrospective interview transcript if available, transcripts of the student stimulated retrospective interviews; and student journal entries about that lesson), folder 5 contained the final test (observation field notes and transcript if available, and relevant documents about the final test) and all the final interviews, folder 6 contained the data from the two questionnaires, and folder 7 contained the teacher baseline interview transcript and all the relevant documents about that site.

As a multiple-case study, within-case analysis was conducted first to gain a detailed understanding of each case before cross-case comparisons were carried out (Miles & Huberman, 1994). During the analyzing stage, data were analyzed in their original language form. During the stage of writing up the thesis, the researcher translated the quotes from interview data and journal data that were originally in Chinese into English and had my colleague who helped with questionnaire translation to check the accuracy of the translation. In what follows, within-case analysis is described first before cross-case analysis.

3.5.1 Within-case analysis

Analysis within one case was carried out in three stages to address the three sets of research questions. The first two stages of analysis: **identifying** and **describing** the identified CA practices, aimed to reveal the typical features of the teacher's actual CA practices, and the focus of analysis was the classroom recording transcripts and observation field notes. Relevant documents and participant interview data relating to the identified CA practices were also analyzed to triangulate the findings obtained from the analysis of the transcripts and field notes. The educational environment in which the teacher was teaching was also derived from the relevant documents and interview data. The third stage of analysis, **revealing** the teacher's explanations and

the students' perceptions of the identified CA practices, aimed to address the second and third research questions, and the focus of analysis was the participants' interview data, and student journal and questionnaire data. Details about each stage of analysis is described below.

3.5.1.1 Identifying CA

To capture a teacher's CA practices, the researcher went through the following three steps. She first analyzed relevant documents such as a course description and teacher interview data to identify those CA practices that were either specified as assessment in the documents or regarded as assessment by the teacher or both. This group of CA practices were regarded as recognized CA practices. An *emic* perspective was embodied in this step of analysis because the participant teachers' ideas of CA were acknowledged. Next, the researcher conducted discourse analysis of the classroom recording data and identified those segments that contained evidence of the three steps of CA: elicitation, judgment-making, and making use of the assessment information. This group of CA practices were regarded as unrecognized CA practices. An *etic* perspective was embodied in this step of analysis because the researcher applied ideas from existing literature to identify CA practices.

In this study, an unrecognized CA practice is identified according to the following two criteria. First, a CA practice should contain evidence of the three key steps of CA: elicitation, interpretation, and using. For example, in the week 14 lesson, Andrew asked his students to do a role play, and after students practiced in pairs for seven minutes, he asked two pairs to perform before the class and gave comments and suggestions after each pair's performance. This activity was regarded as a CA practice because it contained the evidence of the three steps involved in CA. The teacher elicited students' performances by asking students to do the role play, first in pairs and then two students demonstrating before the class; the teacher's feedback, which was contingent upon students' actual performances, reflected that Andrew had evaluated students' performances (the second step), and also made use of the judgment by giving feedback (the third step). Accordingly, when an activity did not contain the step of elicitation, such as when a teacher was giving a lecture on a topic, this activity was not

regarded as a CA practice. Moreover, if an activity had the elicitation step, but there was no sign of the second and third steps of CA, this activity was not regarded as a CA practice either. For example, in Linda's case, she regularly asked her students to give a news report at the beginning of each lesson. However, usually there was no feedback from the teacher or other students after a student's news report. This news report activity was not regarded as a CA practice.

Second, a CA practice usually has clear boundaries. When a CA practice was in the form of a classroom activity, there were always clear discourse markers to mark the beginning and end of a CA practice. As in the above example, Andrew said the following words to mark the beginning of this activity.

> … Now I'm going to spend the rest of the class, we have about 15 minutes, to look at exercise 7. Guys, working in pairs and act out the conversations. Try to learn what you've watched just now and come up with a conversation with you partner. (Andrew_CRT07122010[7]).

And he ended this task by saying "Ok, you know, we don't have enough time for more pairs to come up, but I have this special assignment for you guys" (Andrew_CRT07122010).

When a CA practice took the form of a few turns of classroom interactions, the focus of one CA practice was also different from the focus of other CA practices. For example, when one student was talking about his unlucky weekend during the warming-up chatting of her week 8 lesson, Mary provided online comments/questions and feedback to scaffold the students to finish the task. As can be seen in Excerpt 3.1, there are two CA episodes in this segment of classroom interactions, with turns 1-4 focusing on a pronunciation problem (the pronunciation of the word phone) and turns 9-10 focusing on an expression problem (fly away). Clearly, the two CA episodes had different foci and should be regarded as different CA practices.

Excerpt 3.1

---

[7] In this thesis, the following short forms are used to indicate the source of a piece of data. BI: baseline interview; SRI: stimulated retrospective interview; EoSI: end-of-semester interview; Jack_J1: Jack's journal one; MCR: member checking response; COF 09112010: classroom observation field notes collected on 09 November 2010; CRT28042011: classroom recording transcript of the class on 28 April 2011.

Turn	Speaker	Transcript
1	S1	On Friday night, I went to karaoke with my friends, and when I left I forgot my my full yeah, my full left in Karaoke, [so
2	T	[Forgot your?
3	S2	Cell phone.
4	T	Oh your phone! Oh my God!
5	S1	So I went back. Then when I got got into my dormitory, I realized that, realizing, and I went back to to get my phone.
6	T	Yeah?
7	S1	And when I there is there are stone bricks. I I was riding a bike, and there were stones
8	T	On the road?
9	S1	Narrow stones narrow gap so so I cross to the the gap and my bike and I all threw threw away.
10	T	Really? Went flying, we say went flying.

To enhance the validity of this step in the analysis, one-hour classroom-recording transcripts from each participant teacher's observed lessons were peer-coded by a Chinese colleague, who was and had been an oral English teacher for over 20 years, who had always been one of the most liked teachers in the department, and who had a PhD degree in applied linguistics with a special interest in classroom assessment and language testing. Before being asked to do peer-coding, this colleague was informed of the purpose of this study, the research questions, the broad definition of CA adopted in this study, the various possibilities of CA practices as discussed in section 2.2.2, and the two criteria of identifying a CA practices as described above. While agreement was achieved concerning classroom activities which involved some kind of student performances such as doing a role play, telling a story, having a discussion on a given topic and teacher feedback that was based on students' performances, disagreement mainly lay in whether every initiation-response-feedback (IRF) segments (Sinclair & Coulthard, 1975) should be considered as an FA episode. The researcher initially considered every IRF sequence as an FA episode, but the peer coder pointed out that in some IRF segments there was no evidence to show that the teacher assessed students' performances, as can be seen in the following excerpt.

Excerpt 3.2

Turn	Speaker	Transcript
1	T	er...bonfire, er, you know that, when you go out in a yard or field and build a big fire, everybody sits around, yeah, so that's a bonfire. I think you do that in China sometimes?
2	SS	Yeah, yes.

3	T	When do you do that? When would you have a bonfire in China?
4	S1	When we go camping
5	T	Oh yeah sure, when you go camping. Yeah, we always do that, yeah.

Excerpt 3.2 is taken from Mary's lecture on Halloween (Mary_CR20101026). During her lecture, she raised many questions to students and there were many IRF sequences. In this excerpt, Mary was talking about a bonfire and, after she explained its meaning, she asked her students if Chinese sometimes had bonfires (turn 1). Then turns 3 to 5 were a typical IRF sequence. However, as pointed out by the peer coder, there was no evidence that the teacher evaluated S1's answer. Instead, the teacher was just seeking information from the students. On second thoughts, the researcher agreed with the peer coder and decided that, when both the context and the feedback did not contain evidence of making judgment about students' performances, then an IRF sequence should not be considered as an FA episode. Only when there was explicit evidence to show that the teacher made a judgment about students' performances, could an IRF sequence be considered as an FA episode, as can be seen in the following excerpt.

Excerpt 3.3

Turn	Speaker	Transcript
1	T	Do you get all the answers to the questions?
2	SS	((Some students nodded their heads.))
3	T	Ok, uh, first one, Tae Kwon Do originated about::?
4	SS	2000
5	T	2000 years ago in Korea. Number 2,Tae Kwon Do is basically the::
6	SS	Hand and foot
7	T	Hand and foot martial arts. Number 3, it's open to::
8	SS	((no responses))
9	T	Anyone. We take any flexibility level, any age, and any:: fitness level. Ok, if you are very fit, 很健康对不对((hen jian kang dui bu dui, very fit, right)), if you are kind of overweight, or if you are not flexible, you are not so healthy, you could also go, ok, so any level are welcome. Number 4.

Excerpt 3.3 is taken from the part when Andrew was checking students' answers to a listening comprehension exercise (Andrew_CR20101116). Originally, the peer coder thought Andrew was just telling students the correct answers, regardless of whether they had got them right or not. However, after analyzing Andrew's feedback to every question in this task, the researcher and the peer coder noticed that when most students could get an answer correct (turns 4&6), Andrew would simply repeat as a kind of

confirmation and then move on to the next one (turns 5&7). When most students did not know the answer (turn 8), he would not only provide the correct answer but also give some explanation (turn 9). Then the researcher and the peer coder agreed that Andrew evaluated the class performance on this task and each IRF sequence should be considered as an FA episode. Therefore, during the analysis, when there was not enough evidence to determine if one specific IRF sequence was an FA episode, the researcher would examine the specific context in which the IRF sequence was located to see if there was evidence to prove or disprove it.

3.5.1.2 Describing CA

The identified CA practices were analyzed from two perspectives: cross-sectionally and longitudinally within their courses, to reveal their complexity and dynamics. Cross-sectionally, the analysis started with a preliminary coding list derived from existing relevant literature, and then the researcher went through an iterative process of coding the data and modifying the coding list. During the analysis, the researcher remained open and sensitive to new ideas and categories through continuing "dialogue" between the research questions, the literature and the data (Ritchie & Spencer, 2002). This process entailed reviewing the coded transcripts repeatedly, recoding sections of the data, grouping, merging or removing codes, and identifying illustrative examples. When uncertainty arose, the researcher discussed with her supervisors in detail so that the codes were faithful representations of the data. The final version of the framework for describing a CA practice cross-sectionally is presented in figure 3.3.

A CA practice
- Assessment purpose/function → Why →
  - ➤ Formative
  - ➤ Summative
  - ➤ Washback
- Assessment process →
  - Collecting information → How →
    - ➤ Formal assessment
    - ➤ Planned assessment-involving instructional activity/episode
    - ➤ Incidental assessment episode
  - Making judgment →
    - Who →
      - ➤ Teacher
      - ➤ Students' peers
      - ➤ Students' themselves
    - Construct →
      - ➤ Communicative competence
      - ➤ Idea / content
      - ➤ Learning attitude / behaviour
      - ➤ Other
    - How →
      - Scoring type →
        - ➤ Holistic scoring
        - ➤ Analytical scoring
        - ➤ Intuitive
      - Basis of judgment →
        - ➤ Norm-referenced
        - ➤ Criterion-referenced
        - ➤ Student-referenced
      - Moderation →
        - ➤ Standardization in test administration
  - Making use of the judgment →
    - ➤ Teaching
    - ➤ Learning (feedback)
      - ✓ Person-referenced
      - ✓ Task-referenced
        - ● Confirmative
        - ● Explanatory
        - ● Corrective
    - ➤ Reporting

Figure 3.3 A Framework for Describing CA Cross-sectionally

This framework is built upon the purpose and the three key steps involved in a CA practice. The variables in it are a combination of both variables that have been suggested in some existing researches and those that emerged from this research, as can be seen from the following description of this framework.

The *why* component concerns the purpose /function of a particular CA practice. It has three variables. When students' performances during a CA practice are marked and contribute to the final grading, this practice shows the *summative* purpose. When a CA practice is used to improve teaching and promote learning, it shows the *formative* purpose. When a CA practice is designed particularly to motivate students to work hard, it shows the *washback* purpose, because the teachers wanted the assessment practice to bring about positive learning attitudes or behaviours on the part of the students. While the first two variables were included in Harlen's (2007) framework (cf. Figure 2.2), the third one emerged from the present study. The dotted line in the figure indicates that often the purposes /functions of a CA practice can be inferred from the third step of the assessment process.

81

For the *how* component in the first step—*collecting information*, the present study identified three variables. *Formal assessment* refers to planned and evident CA practices such as tests and marked assignments. *Planned assessment-involving instructional activity/episode* refers to classroom activities or segments of classroom interactions used for teaching by design but for assessment by chance. *Incidental assessment episode* refers to unplanned instruction-embedded segments of classroom interactions that can also be regarded as assessment episodes. The three variables were similar to the three types of CA found in Hill's (2012, pp. 82-83) study but not exactly the same. Hill's term *planned assessment opportunity* refers to an assessment task that is "deliberately embedded in regular classroom activities but learners are not informed that it is an assessment activity" (p.83), but in the present study, the teachers did not regard their instructional activities as assessment. Therefore, the present study adopted a new term. Hill's third variable *incidental assessment opportunity* includes teachers' unplanned observation, which was not examined thoroughly in the present study. The present study mainly focused on segments of classroom interactions that contain assessment, and therefore, a new term *incidental assessment episode* was adopted.

The way to distinguish a *planned assessment-involving instructional activity* from a *planned assessment-involving instructional episode* is that an activity is a bigger unit of analysis than an episode. For example, when Andrew conducted a listening-comprehension-checking activity by asking them to do the listening-comprehension exercises and then provided feedback based on students' responses, the whole activity was regarded as a *planned assessment-involving instructional activity*, for there was evidence of the three steps of an assessment practice. Within this activity, when Andrew asked students one comprehension question and then provided feedback based on students' responses, this segment was regarded as a *planned assessment-involving instructional episode* for it also contained the three steps of assessment. Since the comprehension questions were designed beforehand, this assessment-involving segment was planned in nature.

The way to distinguish a *planned assessment-involving instructional episode* from an *incidental assessment episode* was whether such a segment of classroom interactions is planned in advance or is contingent upon students' actual performances in

classrooms. For example, if a teacher asks a student a question that he/she has planned to ask beforehand, and then after his/her students respond, the teacher provides feedback that helps students to learn, then this episode is a *planned assessment-involving instructional episode*. However, if the teacher's question is derived from students' performances, then the episode is an *incidental assessment episode*. Excerpt 3.3 is an example of this kind, because the teacher's recast in turn 2 was based on the student's response in turn 1.

Excerpt 3.4

Turn	Speaker	Transcript
1	S1	This girl never do exercise and::
2	T	She never <u>does</u> exercise?
3	S1	Yeah yeah yeah. And this girl ((indicating the girl sitting behind him)) almost once a week just like me. And this girl

The second step, *making judgment*, has three components: who is making the judgment, what is the judgment about, and how the judgment is made. For the *who* component, the variable *teacher* means the teacher is the sole agent of judgment, the variable *students' peers* means students are involved in peer-assessment, and the variable *students' themselves* means students are involved in self-assessment. In Harlen's framework (cf. Figure 2.2), the variable *teacher and students together* did not specify whether students are engaged in peer-assessment or self-assessment, which the present study found worth distinguishing. In addition, Harlen's framework contained another variable, *external agent.* However, this third variable did not emerge in the present study and therefore is not included in the present framework.

The *construct* component has four variables. When a CA practice assesses students' oral English ability or its specific aspects, then it is coded with *communicative competence* or its specific aspects. Here Celce-Murcia's (2008) model (cf. Section 2.5, Figure 2.12) served as *a priori* coding list and was found comprehensive enough for the present study. The *content/idea* variable means that the quality of what the students has expressed is assessed, such as the depth, relevance, or logic of their ideas. The *learning behaviour/ attitude* variable refers to such non-academic aspects as class attendance, student learning attitude, motivation or interest. This study also revealed a few constructs that falls outside the above three categories and therefore grouped

83

under the *other* variable. For example, Andrew assessed students' cooperation skills, Andrew's final test assessed students' memorization, and the three teachers all assessed students' use of visual aids.

The *how* component is mainly concerned with formal assessment practices because this component concerns grading and reporting. It has three sub-components: *scoring type*, *basis of judgment*, and *moderation*. For *scoring type*, two variables emerged from the present study. *Analytical scoring* means teachers give marks for different aspects of students' performances, and *holistic scoring* means teachers give one mark for overall impression. These two variables are the same as those in Cheng and Wang's (2007) study, but they had a third type of scoring: rubric scoring, which was not found in the present study.

For *basis of judgment*, three variables appeared in the present study. The *norm-referenced* variable means a teacher's judgment of a student's performance is made by comparing this student's performance against other students' performances and then ranking this student in relation to his/her peers. The *criterion-referenced* variable means a teacher's judgment of a student's performance is made by measuring the student's performance against a fixed set of predetermined criteria or learning standards. The *student-referenced* variable means a teacher's judgment of a student's performance is made by comparing this student's performance with his/her previous performances. These three variables were similar to the variables in Harlen's model (cf. Figure 2.12), though Harlen divided criterion-referenced into two types: task-specific or non-task-specific.

Moderation is a process used to prevent errors or bias in the process of an assessment (quality assurance) or in the outcome of an assessment (quality control) (Harlen, 2007, pp.76-77). It is concerned with the reliability of teacher-made assessments, especially for summative purposes. Usually the higher the stakes of an assessment, the more important such moderation processes are. While Harlen (2007, pp.76-79) identified several moderation strategies, the present study only found one instance of moderation. That was in Linda's final test where a certain level of standardization in test administration was achieved to assure test quality. Therefore, *standardization in test*

administration is the only variable for the *moderation* component in the present framework.

The third step, *making use of the judgment*, consisted of three variables. *Teaching* means the judgment of students' performances is used to inform teaching; *learning* means such judgment is used to promote student learning, which is realized through *feedback* to students; and *reporting* means such judgment is used to inform decisions about end-of-semester grading and reporting. The above-mentioned variables are the same as those listed in Hill and McNamara's (2012) framework (cf. Table 2.1), although they had two more variables, *classroom management* and *socialization*, which did not emerge in the present study.

Regarding types of feedback to students, this study found the classification proposed in Hill and McNamara's (2012) framework clear, comprehensive, and examples of each type were found from the present study. They classify feedback into two major types: *person-referenced feedback* that focuses primarily on the student's ego, and *task-referenced feedback* that focuses primarily on the student's performance in relation to specific aspects of task requirements and/or qualities and standard of performance. Task-referenced feedback is further classified into three types: *Confirmative feedback* to acknowledge students' answers/ responses as correct, good, relevant, or significant; *explanatory feedback* to provide specific information about their performance in relation to task specifications and/or qualities and standard of performance; and *corrective feedback* to show students the gap between what is expected and students' actual performances by providing true answers (Hill & McNamara, 2012).

This framework was used to describe and generate profiles of the identified CA practices in each case. In addition to the cross-sectional description of the identified CA practices, the researcher also paid attention to when the identified CA practices occurred, how deeply they were embedded within instruction, and how different CA practices were related to each other. This kind of analysis was to reveal the dynamics of CA practices in context.

3.5.1.3 Revealing the participants' perceptions

While the above two steps of analysis were used to address the first set of research questions, for the second and third sets of research questions, that is to reveal the teacher's opinions and the students' perceptions of the identified CA practices, qualitative content analysis (Dornyei, 2007, pp. 245-257) was conducted on the interview and journal data. These data were read and reread to identify the parts relating to specific CA practices or relating to the participants' beliefs about CA in general. The intention was to let the data speak so as to reveal the participants' beliefs, feelings, and perceptions (Ellis & Barkhuizen, 2005; Mackey & Gass, 2005).

Finally, student questionnaire data from each class were analyzed independently using SPSS and the findings were triangulated with the findings from the qualitative data. Specifically, the beginning questionnaire data went through descriptive analysis to reveal students' characteristics before the start of the study. Regarding the end-of-semester questionnaire, for Andrew's case and Mary's case, the second part of the questionnaire was analyzed to see if certain changes had occurred through one semester of study concerning students' oral English learning. Items were first grouped into six aspects: students' self-evaluation (items 1 to 7 with items 2 and 4 reversed), mastery goals (items 8, 12, 15), performance approach goals (items 9, 11, 14), performance avoidance goals (items 10, 13, 16), anxiety (items 17-19), and effort (items 20-23). Then an independent sample t-test was conducted on each aspect with the beginning data as time 1 and end-of-semester data as time 2. As to the third part, originally it was planned that descriptive statistics should be calculated and then two Pearson correlation analyses should be carried out: one between frequency of the selected practices and the six aspects relating to students' oral English learning in part II of the questionnaire and one between the usefulness of the selected practices and the six aspects of students' oral English learning in part II. However, because the questionnaire turned out to be problematic (cf. Table 4.13), such analyses were omitted. For Linda's case, descriptive analysis for each section of the end-of-semester questionnaire data was conducted.

3.5.2 Cross-case analysis

Following the suggestion that it is important to gain an understanding of the complexity and dynamics of each case in its own right before attempting to make cross-case comparisons (Miles & Huberman, 1994), in this study, cross-case analysis was not conducted until an intimate knowledge of each teacher's CA practices was achieved and the profiles of each teacher's CA practices were member checked by the participants. This stage of analysis was guided by the descriptive framework for CA for this study (cf., Figure 3.3). By comparing and contrasting the findings concerning each aspect of CA from the three cases, the commonalities as well as distinctive features of the three teachers' CA practices were synthesized, the overriding themes governing the teachers' CA practices were extracted, and the effective CA practices were revealed. During this comparative analysis, whenever an uncertainty arose, the researcher would go back to the original data to check or discuss it with the supervisors. Then the identified features were analyzed in relation to existing literature. The results of such comparative analysis are presented in chapter 7.

3.6 Summary

This chapter presented the rationale for the research design and described how data were collected and analyzed for this study. The following three chapters will present results from each of the three cases respectively. The results of the comparative analysis will be presented in chapter 7.

Chapter 4. Andrew's CA Practices

This chapter describes the results of the data analysis for Andrew's case. After the description of his teaching context, Andrew's formal assessment structure is presented first. Then the profiles of the components within his formal assessment structure are presented separately before the profile of his unrecognized CA practices are described. Each profile is presented following the framework presented in figure 3.3. Students' general impressions of his CA practices as revealed from the students' questionnaire data are presented after that. This chapter ends with a summary of the typical features of his CA practices. The next two chapters are about the other two teachers' CA practices, and they follow the same presentation structure.

4.1 Andrew's teaching context

4.1.1 Educational background and past teaching experiences

Andrew, a Chinese male in his early 30's, obtained his B.A. degree in English Education and an M.A. degree in American Literature consecutively from two universities in the northwest part of China. Upon graduation, he became an English teacher teaching English language skill courses to adult learners at the Continuing Education College of his university, a key foreign language university in Beijing. At the time of the present study, he had just moved to the School of English for Specific Purposes (ESP) of his university but he had taught at his university for seven years. The present study was conducted in the fall semester of 2010, his first semester as a teacher at the ESP school, where he taught oral English and English writing courses to non-English-major undergraduates. Though it was his first time teaching this oral English course to undergraduate students, it was not the first time he had taught oral English.

Andrew had little training in assessment. During his undergraduate study, he had learned some English teaching methodology, but had no specific course on language testing or assessment. Immediately after he became a teacher, to obtain his teacher's certificate, he received the required teacher training which focused on educational

psychology and teaching methodology where assessment was touched upon only slightly (Andrew_BI).

4.1.2 His university, his school, and his students

The university where Andrew was teaching specializes in foreign language teaching and offers the most number of foreign language programs in China. The ESP school was founded in 2008, the newest of the 14 schools of his university. The School is responsible for offering English courses, both compulsory and optional, to all non-English-major students at both undergraduate and postgraduate levels. Information from the university website shows that English learning at this school is divided into eight bands. Bands 1 and 2 are for those non-English-major students who have never learned English formally before they come to university. These students are mainly from foreign language schools and have studied Japanese or Russian or a language other than English in their middle school and high school. For these two bands, only one course—Intensive Reading—is offered each semester. Students are expected to complete the two bands in four semesters and they can earn 16 credits altogether. Bands 3 and 4 are for those non-English-major students who have learned English in their middle school and high school years. Each band offers three courses: Intensive Reading (4 credits), Oral English (2 credits), and Multimedia Listening (2 credits). Courses for Bands 1-4 are free for students, and the university requires that each non-English-major student should earn 16 credits from English courses before they are allowed to graduate. Courses for Bands 5-8 are optional and students have to pay if they want to study those courses. If a student has successfully earned 16 credits from the courses for Bands 1-4 and earned all the 32 credits from the courses for Bands 5-8, he/she can get a diploma in English, which is a second qualification because students are all working towards a degree with a major other than English (from the university website, accessed on Feb. 17, 2012).

The present study was conducted in the Band 3 Oral English course. Andrew was teaching this course to two of 20 parallel classes. Data from the student beginning-of-semester questionnaire showed that the class Andrew chose for the study had 29 students altogether, 17 male, 12 female, majoring in Japanese, German, or

Italian. Only two students were from foreign-language high schools where they had learned English intensively but they had decided to study a different language at university. Except for four students who started to learn English from primary school, the majority of the students had started to learn English since becoming middle school students. Four students had travelled abroad for a period from ten days to one month.

Questionnaire data also revealed that at the beginning of the semester, students did not think very highly of their own oral English abilities (Table 4.1). Half of the students felt a sense of inferiority to those who could speak fluent English, and more than two thirds of the students were not satisfied with their present oral English proficiency. In addition, nearly half of the students felt their oral English was not very fluent and they sometimes could not express themselves clearly.

Table 4.1 Andrew's Students' Self-Evaluation of their Oral English Ability at the Beginning of the Semester

Questionnaire item	Not true at all	Not really true	Partly true	Mostly true	Absolutely true	Mean
1. I am very confident when I speak English.	1 (3.4%)	3 (10.3%)	17 (58.6%)	7 (24.1%)	1 (3.4%)	3.14
2. I worry that native English speakers will find my oral English strange.	4 (13.8%)	6 (20.7%)	8 (27.6%)	9 (31%)	2 (6.9%)	3.03
3. My oral English is very fluent.	4 (13.8%)	8 (27.6%)	13 (44.8%)	3 (10.3%)	1 (3.4%)	2.62
4. My poor oral English always makes me feel inferior to those who can speak fluent oral English.	0	7 (24.1%)	7 (24.1%)	7 (24.1%)	8 (27.6%)	3.55
5. When I speak in English, I can always find the words to express my ideas.	2 (6.9%)	10 (34.5%)	11 (37.9%)	5 (17.2%)	1 (3.44%)	2.76
6. When I speak in English, I can say exactly what I want to say.	1 (3.4%)	11 (37.9%)	13 (44.8%)	4 (13.8%)	0	2.69
7. I am satisfied with my present oral English proficiency.	7 (24.1%)	15 (51.7%)	3 (10.3%)	3 (10.3%)	1 (3.4%)	2.17

Moreover, at the beginning of the semester, the students' goal orientations were mainly towards mastery of the learning objectives (mastery goals) and doing well in class and exams (performance approach goals) rather than avoiding punishment or failure (performance avoidance goals), as can be seen in Table 4.2.

Table 4.2 Andrew's Students' Goal Orientations at the Beginning of the Semester

Questionnaire item	Not true at all	Not really true	Partly true	Mostly true	Absolutely true	Mean
Mastery goals						
8. It's very important for me to completely grasp the skills taught in this course.	0	1 (3.4%)	4 (13.8%)	10 (34.5%)	13 (44.8%)	4.25
12. I am willing to spend a lot of time practicing to improve my oral English.	0	3 (10.3%)	8 (27.6%)	9 (31%)	9 (31%)	3.83
15. I believe that if I work hard, my oral English will improve.	0	0	4 (13.8%)	9 (31%)	16 (55.2%)	4.41
Performance approach goals						
9. It is very important for me to be the top student in this oral English class.	0	7 (24.1%)	11 (37.9%)	6 (20.7%)	5 (17.2%)	3.31
11. I am eager to become a smart student in my teacher's eye in this course.	0	2 (6.9%)	6 (20.7%)	12 (41.4%)	9 (31%)	3.97
14. I wish to get a high score in this course.	0	1 (3.4%)	5 (17.2%)	10 (34.5%)	12 (41.4%)	4.18
Performance avoidance goals						
10. I will be satisfied so long as I can pass this course.	6 (20.7%)	9 (31%)	9 (31%)	5 (17.2%)	0	2.45
13. I worry that I may fail this course.	7 (24.1%)	13 (44.8%)	2 (6.9%)	5 (17.2%)	2 (6.9%)	2.38
16. I feel that no matter how hard I try, my oral English will make little improvement.	12 (41.4%)	13 (44.8%)	3 (10.3%)	1 (3.4%)	0	1.76

In addition, students' anxiety levels varied but in general were not very high (Table 4.3). Similarly, their enthusiasm to improve their oral English was not very high either (Table 4.4).

Table 4.3 Andrew's Students' Anxiety Level at the Beginning of the Semester

Questionnaire item	Not true at all	Not really true	Partly true	Mostly true	Absolutely true	Mean
17. I get very nervous when a foreigner asks me something in English.	3 (10.3%)	4 (13.8%)	9 (31%)	7 (24.1%)	6 (20.7%)	3.31
18. I feel distressed about being unable to improve my oral English.	8 (27.6%)	11 (37.9%)	6 (20.7%)	2 (6.9%)	2 (6.9%)	2.28
19. I worry that other students will laugh at me when I speak English.	4 (13.8%)	13 (44.8%)	7 (24.1%)	4 (13.8%)	1 (3.4%)	2.48

Table 4.4 Andrew's Students' Effort at the Beginning of the Semester

Questionnaire item	Not true at all	Not really true	Partly true	Mostly true	Absolutely true	Mean
20. I actively create opportunities to talk with others in English.	1 (3.4%)	14 (48.3%)	10 (34.5%)	4 (13.8%)	0	2.59
21. I regularly enlarge my oral English vocabulary.	1 (3.4%)	10 (34.5%)	14 (48.3%)	2 (6.9%)	2 (6.9%)	2.79
22. I try my best to improve the grammatical accuracy of my oral English.	0	6 (20.7%)	12 (41.4%)	5 (17.2%)	5 (17.2)	3.32
23. I grasp every opportunity to practice my oral English.	3 (10.5%)	10 (34.5%)	11 (37.9%)	5 (17.2%)	0	2.62

4.1.3 The oral English course

Andrew's Band 3 Oral English course was a 16-week course, 2 hours per week. Before the semester started, Andrew was given the course description (Appendix 12), in which this course was described as focusing on "topics related to daily life and social issues", and students were expected to make improvement in "fluency", "accuracy", "vocabulary", "English culture", and "self-confidence in expressing themselves in English," through various activities such as "conversations, story-telling, presentations."

The required textbook was *New Standard College English (Level 2): Listening and Speaking* (Greenall, Tomalin, & Friedland, 2008). While the textbook includes ten units, only seven units were covered during the semester, since each unit required four hours. Each unit in the textbook, following the same pattern, contained nine sections (cf. Appendix 13). Classroom observation showed that Andrew followed the course syllabus strictly, and as for each unit, he generally followed the textbook though sometimes he added or omitted certain activities in the textbook to "achieve better effects" (Andrew_SRI).

In Andrew's mind, the course objectives as specified in the course description seemed a little "vague," and he thought "since this book has been chosen, we should expect students to learn something from it. … Since each unit has its specific learning

objectives, I just set these specific objectives as the learning objectives for my students" (Andrew_EoSI).

Moreover, Andrew joined this teaching group for the Band 3 Oral English course at the last minute. "One evening I got a phone call telling me that I would teach oral English Band 3 and my class was scheduled the next morning" (Andrew_BI). Because his email address was not included in the group list, he was often late in receiving messages from the group. For example, for the final test, on the night before his week 15 lesson, Andrew realized that the next day he had to tell his students about the final exam, but he had not received any information about it at that stage, so he phoned the course leader who then emailed him the relevant information (Andrew_EoSI). Such poor communication and his first-time teaching made him not very clear about the course objectives or the student assessment, and consequently he had to "feel my own way out" (Andrew_EoSI).

As Andrew was given a syllabus, a textbook, as well as the examination instructions and methods, his working environment was towards the "high structure" end, where "teachers are obliged to follow a comprehensive, pre-specified syllabus as well as a textbook and/or examination prescription, all of which may have been developed externally. Learners have no, or very few, curriculum responsibilities" (Wette & Barkhuizen, 2009, p. 198).

This course was important for the students who took it, because if they failed it they had to take it again until they passed it, or they might not get their BA degree. On the other hand, the course was not that important for the students in the sense that it would not affect their comprehensive evaluation at the end of the semester, nor would it play a role in determining whether a student could win a scholarship. In other words, it would make no difference if a student just barely passed the course or did so with distinction.

4.2 Andrew's CA practices

4.2.1 Formal assessment structure

According to the course description (Appendix 12), students' performances in the following four aspects contributed to their final composite scores of this course: *class attendance* (10%), *class participation* (30%), *presentation* (10%), and *exam* (50%). While *class attendance* is easy to understand and the course description provided details about the *presentation* task and the *exam*, the *class participation* aspect remains vague in the document.

Teacher interview data showed that in Andrew's mind, he mainly used the *presentation* task and the *exam* to assess students in this course. During the end-of-the-semester interview, Andrew said the *presentation* task accounted for 10% of the final score and the *exam* 80%, and he regarded *class participation* as more or less the same as *class attendance*, which together "take up 10% of the final grading" (Andrew_EoSI). He added that "*class participation* was meaningful in the final grading only in the sense that it could reflect students' exposure to certain teaching materials, but in fact should not be taken into consideration in the final grading because it was not related to students' oral English ability" (Andrew_EoSI). However, in his member checking response a few months later, Andrew corrected himself saying that "I didn't remember the percentages for each part correctly during the interview, but when I later calculated students' final composite scores, I did follow the percentages as indicated in the course description" (Andrew_MCR).

As a new teacher, Andrew felt he had to comply with the assessment instructions specified in the course description and those given by the course leader. For class attendance, classroom observation showed that during each lesson Andrew called out the students' name list to check their attendance. For the presentation task, "our group leader said that so long as students can complete their presentations, they should all get full points for this part" (Andrew_SRI). Since there was no clear instruction as to how to assess student participation, Andrew worked out the following way to both comply with the course requirement and to assess students in a comparatively fair and objective way: On the one hand, he gave most students full scores for class attendance

and for the presentation task. On the other hand, he used students' presentation scores to represent their class participation scores.

> I probably won't deduct many points in this presentation section. I probably will deduct one to two points at most. But my records of students' performances will be reflected in the participation section, or my general impression scores. In this way, I can sort of differentiate the students in this class according to their real oral English abilities. (Andrew_EoSI)

In sum, Andrew's formal assessment structure consisted of three parts: student presentation task, a final exam, and student class attendance. Figure 4.1 summarizes these features.

| Wk 1 | | Wk 10&11 | Wk 14&15 | Wk 16 |

Attendance

Class participation (based on students' presentation performances)

Final test

Completion of the presentation task

Final score   10%   30%   10%   50%
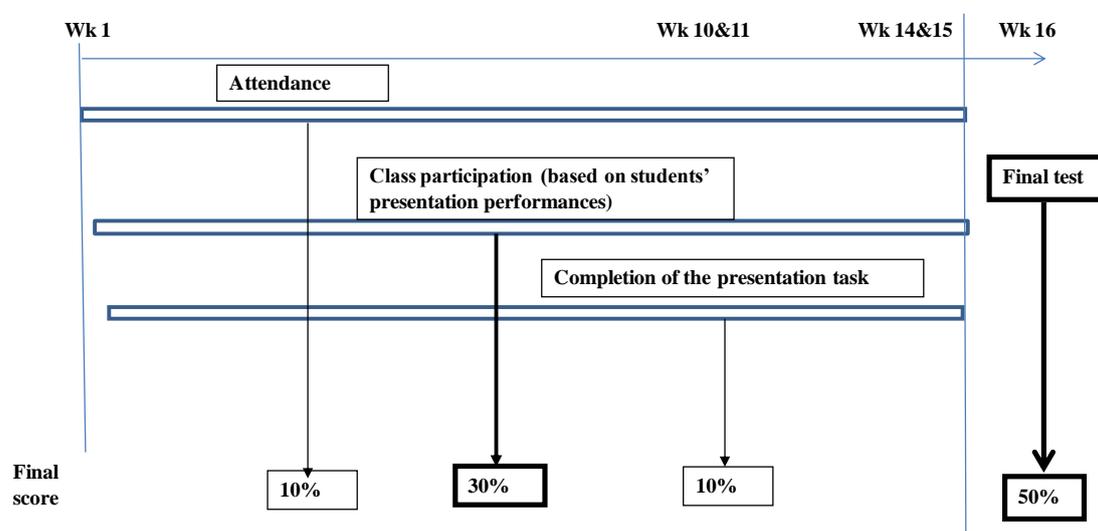
Figure 4.1 The Formal Assessment Structure in Andrew's Oral English Course

The above assessment structure revealed that the students' final scores were not solely about their academic achievement, but also to some extent reflected students' learning behaviours or learning attitudes as indicated by their class attendance. Moreover, classroom observation and analysis of classroom recording transcripts showed that, apart from the presentation task and the final test, there were also some assessment practices embedded within Andrew's daily classroom teaching, though Andrew did not regard those practices as assessment. In what follows, the profiles of the *presentation* task and the *final exam* will be presented before the description of those unrecognized CA practices.

4.2.2 The presentation task

To generate a profile of the presentation task, the author analyzed the general guidelines about this task as specified in the course description (Appendix 12), the classroom recording data and the observation field notes relating to the presentations during the four observed lessons, as well as the relevant parts in students' journals, interviews and the teacher interviews. Using the framework outlined in chapter 3 (Figure 3.3), this presentation task is described from the following four aspects: assessment purpose, information-collecting method, judgment-making process, and making use of the judgment. Students' perceptions about this task are described after the profile of this task.

4.2.2.1 Assessment purpose

According to Andrew's course description (Appendix 12), during the semester each student should give a presentation on a theme related to the course for about 2-3 minutes. "The presentation is a practice of public speaking and teamwork, and will account for 10% of the course assessment." This instruction indicates that this assessment task was used for dual purposes: both as a practice opportunity, which should help students learn and improve their presentation skills, thus formative in nature, and as a way to generate a score for the grading and reporting purpose, thus summative in nature.

Furthermore, an additional purpose of conducting this task was to have positive washback effects on students. The interview data revealed that this presentation task was originally designed more as a practice opportunity than as an assessment task. "Our group leader said that so long as students can complete their presentations, they should all get full points for this part" (Andrew_SRI). Since every student who finished this task would get a full mark, students' marks would not reflect their presentation skills. However, making this task account for 10% in the final grading "would make students take this task more seriously" (Andrew_SRI). Therefore, the marking element attached to this task was made a tool to make students work harder, a typical assessment purpose of Chinese EFL teachers as found in Cheng, Rogers, and Wang's study (2008).

4.2.2.2 Information-collecting method

This presentation task belonged to the category of *formal assessment*. For one thing, the presentation task was not only formally specified in the course description (Appendix 12), which was given to the students at the beginning of the semester, but was also fully recognized by both Andrew and his students. Classroom observation showed that during the observed lessons, all the presenters appeared well prepared, as seen from their carefully prepared scripts and the well-made power point slides. Some even inserted video clips or background music in their presentations. When several students were presenting together, they paid attention to the transition from one presentation to the next (Andrew_COF09112010). During the end-of-semester interviews, all the volunteer students regarded the presentation task as assessment. For another, classroom observation showed that the presentation task, as a routine task at the beginning of each lesson, to some extent stood apart from Andrew's teaching agenda. This kind of separateness can be seen from the low connection between the students' presentation topics and what Andrew was about to teach on a particular day (see Table 4.5), which made this task appear more test-like.

Table 4.5 Information about Student Presentations in Andrew's Class

Week	Presentation topic	Students	Form and total time	Peer feedback	Total feedback time
10 (Unit 6: Sporting Life)	Sports: Tennis, basketball, football (related to Unit 6: Sporting Life)	3 female students	S1&S2 Dialogue → S1 presentation → S2 presentation → S2&S3 dialogue → S3 presentation → S1 summary (7 min)	Yes (5 students made 5 very short comments)	6 minutes (mainly from Andrew)
11 (Unit 6: Sporting Life)	Politicians: David Cameron & G. W. Bush (related to Unit 4: News 24/7)	2 male students	S1 presentation → S2 presentation (18 min)	Yes (2 students made 3 very short comments)	7 minutes (mainly from Andrew)
14 (Unit 9: Job Fair)	TV series: Lie to Me (related to Unit 2: Crime Watch)	1 female student	Her presentation (15 min)	No	6 minutes (all from Andrew)
15 (Unit 9: Job Fair)	Historical relations between Germany and Italy (related to Unit 5: the World at War)	3 male students	S1 introduction → S2 presentation on Germany → S3 presentation on Italy → S1 summary (15 min)	No	6 minutes (all from Andrew)

4.2.2.3 Judgment-making process

Who

Regarding the agent of judgment, classroom observation showed that Andrew sometimes involved students in peer assessment before he provided his own feedback, but not to a great extent. He was always the major agent of judgment (see Table 4.5). For example, in the lessons of week 10 and week 11, after the group presentations, Andrew asked students to comment first. However, only a few students made some simple comments like "great topic", "the music is a bit loud", "maybe more interactions", and "I think when he delivered the speech he turned back to the screen". After each comment, Andrew elaborated on it extensively and added a few more points before he summarized the key points for each presenter. In the lessons of week 14 and week 15, students were not involved in commenting on their fellow classmates' presentations and Andrew gave long and detailed feedback to each presenter (see Table 4.5). Obviously, students were only slightly involved in peer-assessment and there was no sign of self-assessment.

Construct

Analysis of the marking criteria for this task (see Appendix 12) revealed that the task was supposed to assess students' communicative competence, especially their linguistic competence, discourse competence, and socio-cultural competence. However, the last point in the marking criteria, how well students cooperate with their partners, was beyond an individual student's communicative competence and actually concerned cooperation skills when a student worked together with other students in a presentation (see Table 4.6).

Analysis of the feedback session after each presentation during the observed lessons showed that Andrew looked at students' linguistic competence, discourse competence, and paralinguistic competence, which are part of a student's communicative competence; but he also looked at students' cooperation skills when they gave group presentations as well as their use of visual / aural aids. Table 4.7 summarizes the aspects mentioned in the feedback sections of the observed presentations, the

frequency of each aspect (in brackets), and an example of each aspect. It can be seen that Andrew put more emphasis on paralinguistic and linguistic competence than on discourse competence. In addition, as specified in the marking criteria, he commented on students' cooperation skills. In addition, he looked at students' use of visual / aural aids, which, though not included in the marking criteria, he considered important and the effective use of those aids needed to be made known to the students (Andrew_SRI).

Table 4.6 Constructs of Andrew's Presentation Task as Revealed through the Marking Criteria

Marking criteria	Assessment constructs	Categories of English Speaking Ability
Whether the students speak <u>fluently</u> and with <u>correct pronunciation</u> and <u>grammar</u>;	Speaking fluency Pronunciation Grammar accuracy	Delivery → Linguistic competence
whether they possess an <u>appropriate and effective variety</u> of vocabulary and can express their ideas in <u>flexible sentence structures</u>	Vocabulary diversity Vocabulary precision Grammar range	Language use → Linguistic competence
Whether the ideas are conveyed in an organized way and cover the necessary aspects of the given topic	Coherence Idea progress Content relevance	Topic development — Discourse competence
Whether the students speak with <u>proper manners</u>	Polite behavior in social situations	Sociocultural competence
and cooperate with their partners well	Team work	Cooperation skills

Table 4.7 Aspects Mentioned during Andrew's Feedback on Students' Presentations

Aspect	Example
Paralinguistic competence (12)	
Body language (9)	And the first speaker, I mean all of them talked to the audience, but there was this lack of eye contact. Ok, you guys didn't look at the audience very much. You were either looking at the picture or you were looking at the table or the thing. You didn't look at the audience, but the second speaker actually had some eye contact.
Voice volume (3)	So the music was too loud and your voice was a bit too low, so you didn't actually project. It's said in presentations we always project to the audience, right.
Linguistic competence (9)	
Pronunciation (3)	And the last speaker, you have to pay attention to, you lose some sounds,

	you know, in a word, for example, "abou" , "about" , "Importan", "t" is lost, ok.
Vocabulary (3)	T: […] most importantly, the second speaker used very good words, kind of professional. For instance, economic recession, 经济萧条(Jing Ji Xiao Tiao, economic recession), right? Uh, 比如说经济呢突然，它的popularity 突然下降怎么说？ ((For example, how do you say "its popularity dramatically went down"?)) S1: Dramatically. T: Yeah, dramatically went down. Ok, this is a very good phrase. And …, 在他的总统任职期间有两次经济萧条，怎么说？ ((How do you say "there are two economic downturns in his presidency"?)) Did you guys get that? Ss: (xxx) T: Two economic downturns in his presidency. These are very very good words. Two economic downturns in his presidency, ok.
Speaking speed / fluency (2)	I guess one point, ok, she might need to pay more attention to, is the speed of speech. Ok, you have to make sure that the purpose of your presentation is to get other people understand what you are saying, to get the information across. If you speak too fast, or you know unclearly, that might influence the result. So just pay attention to that, ok.
Grammar (1)	T: Oh, this one, so the birthday of David Cameron, so when is his birthday? Ss: (xxx) T: It's on October the 9th. Ok, you don't say October 9, you say October the 9th. ok.
Discourse competence (1)	
Idea progression (1)	First of all, she started with an introduction. She said I'm going to talk about something that is closely related to the course, and in the very end, she made some comments, she gave us a conclusion. So we have an introduction, and then a conclusion. And the whole presentation is very well organized. she said she has 4 parts, so every time when she finished one part, she pointed that out to us, so we know that this is a very well organized presentation: introduction, main characters, key words, and some classics.
Cooperation skills (4)	
Team work (4)	Team work, ok, very very good coordination and team work. I like one speaker is talking here, the other will be controlling the power point and they will also change the music. Ok, that's good.
Use of aural / visual aids (5)	
Aural aids (3)	Yeah, the music was a bit loud, it was too loud. I was standing here, I mean it's good to have some background music, but I was standing here, and I didn't have time you know catch what they said. So it was too loud and it was very distracting.
Visual aids (2)	At the very beginning, she showed a trailer, we call that a trailer, a video clip, ok, 一个预告片((yi ge yu gao pian, a trailer)). That trailer all of a sudden grabbed our attention. I was very interested in that trailer, and I might look for that show to have a look.

Clearly, Andrew focused on more aspects than those specified in the marking criteria and he seemed to emphasize paralinguistic competence more than discourse competence and socio-cultural competence.

How

Regarding the scoring type, the interview data showed that Andrew practiced holistic scoring. Andrew said that when evaluating students' presentations, he mainly looked at "students' overall performances during the presentation," including their "language quality" as well as their "presentation skills" such as their body language, and then gave an impression score for each student (Andrew_SRI). Though he mentioned some specific aspects, he did not give separate marks for each aspect, and therefore his scoring was holistic in nature.

Regarding the basis of judgment, evidence showed that his basis of judgment was not NRM. Classroom observation revealed that while Andrew always provided detailed feedback after each presentation, he never compared one student's performance against another's. There was some evidence to show that his judgment-making process was criterion-referenced, because he was given the marking criteria (Appendix 12) and he did consider several aspects before he reached a conclusive mark for a student's performance, as mentioned above. However, later Andrew explained that his deciding on a mark for a student's performance mainly depended on his teaching experience and his understanding of what a good presentation should be like because there were no specific standards for him to refer to (Andrew_MCR). This reflected the fact that his judgment was more intuitive in nature.

There was no evidence that his marks on students' presentations were moderated in any way.

4.2.2.4 Making use of the judgment

Analysis showed that Andrew's evaluation of students' presentations was used in two ways. On the one hand, his evaluation of one student's presentation contributed to the final grading of that student. As mentioned in section 4.2.1, Andrew did score students' presentations and this part was worth 30% in the final grade. Therefore, his judgment was used for reporting purposes.

On the other hand, his judgment was used formatively to help with student learning. Classroom observation showed that after each presentation, Andrew always provided long and detailed feedback to the whole class based on the specific students' performances (see Table 4.5), and he explained that

> because we have a lot of things to cover for each class, I cannot teach students this 'Presentation Skills' section for every unit. So my usual way is asking them to do the presentations first and then giving them feedback on the skills that are frequently used. Through this kind of repeated emphasis on these skills, they will gradually become aware of them. You know, very often I ask students to give comments first, … and then I will summarize all the important points. (Andrew_SRI)

He also hoped that "through this kind of repeated emphasis on these skills, they [students] will gradually become aware of them" (Andrew_SRI). Obviously, Andrew's judgment was used to help with student learning.

This formative function was further strengthened by the way Andrew downplayed the grading purpose of this task. Although Andrew took careful notes about each student's performance on the task and gave each student a mark, he never told students their scores, because

> I want students to realize that scores are not the most important thing. The most important thing is that they grasp those presentation skills by way of giving their own presentations and taking in my feedback. You know, once you told them the scores, they would care about the scores very much, … and then their study focus would change. So I want them to put aside the scores and focus on how to improve their language ability and how to give presentations. (Andrew_SRI)

Further analysis of classroom recording field notes and transcripts revealed that his feedback was all task-referenced and was primarily explanatory (24 instances altogether), and when he provided feedback on students' vocabulary / grammar / pronunciation errors, his feedback was corrective (7 instances altogether).

4.2.2.5 Profile of the presentation task

All the above-mentioned features of the presentation task are presented in Figure 4.2. The numbers after specific types of feedback show the frequencies of respective types of feedback found in the data.

Why
- ➢Formative
- ➢Summative
- ➢Washback

Collecting information → How → ➢Formal assessment

Who → ➢Teacher / ➢Students' peers

Andrew's presentation task

Making judgment

Construct →
- ➢Communicative competence
  - ➢Interactional competence
    - ✓**Paralinguistic** competence
  - ➢**Linguistic competence**
  - ➢Discourse competence
- ➢Other:
  - ➢Cooperation skills
  - ➢Use of visual / aural aids

How →
- Scoring type → ➢Holistic scoring
- Basis of judgment → ➢Intuitive

Making use of the judgment →
- ➢**Learning (feedback)**
  - ✓Task-referenced
    - ●**Explanatory (24)**
    - ●Corrective (7)
- ➢Reporting

Figure 4.2 Profile of Andrew's Presentation Task

4.2.2.6 Students' perceptions

Five students volunteered to participate in the journal writing and student interviews. Their profiles can be found in Table 4.8. The last dimension, the determination of students' oral English ability, was based on the researcher's observation of students' performances in class and Andrew's judgment during teacher interviews. Students were roughly put into one of three categories according to their positions in relation to other students in their own class. The oral abilities of the volunteer students in Mary's and Linda's classes were determined in a similar way.

Table 4.8 Background Information on the Volunteer Students in Andrew's Class

Student psydonym	Gender	English-learning starting age	High school type	Going-abroad experience	Oral English ability
Jack	Male	13-15	Non-FL[8]	No	Average
Tom	Male	13-15	Non-FL	No	Low

[8] FL stands for foreign language high school. In China, there are a few foreign language high schools where students can have the chance to study a foreign language such as English, Japanese, Russian, etc., in a much more systematic way than those who study in a non-foreign language high school.

Tina	Female	9-12	Non-FL	No	Average
Nancy	Female	13-15	Non-FL	No	Low
Jenny	Female	9-12	Non-FL	No	High

In general, the five volunteer students all found this task beneficial, mainly in two ways.

On the one hand, they found the task helpful in improving their presentation skills, especially their paralinguistic competence such as projecting one's voice, speaking in an unaffected manner, wearing a smile, using proper body language, maintaining good eye contact with the audience, etc.. For example, for the two students' presentations in the week 11 lesson, Jenny wrote in her journal that,

> Obviously they had made careful preparations for this task since they had such long scripts. Their learning attitude is really very good. But the shortcoming is that the first student just read aloud his script and didn't actually communicate with the audience. Also his speed was a little too fast and I didn't hear him clearly. In addition, he turned his back to the audience when he gave his presentation, which indicated that he was not very polite and not very confident. The other student's topic was very interesting. The video clip really attracted the audience's attention. His presentation was very humorous and attractive, full of very good sentences. However, his body language revealed his nervousness, and he also spoke too fast. I will learn from their good points and avoid their shortcomings in future. (Jenny_J2)

On the other hand, they felt this task helped broaden their world knowledge. For example, after the two boys' presentations on the two presidents, Jack wrote in his journal that "the two students' presentations were very interesting, especially the second student's presentation on George W. Bush. His introduction and his comments were very interesting, and I have learned some interesting anecdotes about President Bush" (Jack_J2). Tina wrote in her journal that "I'm not a person who is interested in politics, but their presentations helped me see the less serious side of politics. I think I'll try to read more of the current political affairs in the US or the UK and pay more attention to politics" (Tina_J2). After the presentation on the TV series *Lie to Me*, Tina wrote in her journal that "from today's presentation, I have learned that people's different facial expressions mean different things, so we can judge what a person is thinking from what he/she looks. I find this kind of psychology very interesting" (Tina_J3). Therefore, this task helped some students increase their world knowledge.

Moreover, data also showed that while this task created little anxiety for the two comparatively more competent students, it caused a lot of anxiety for the two struggling students. Nancy and Tom, who did not seem to have a good experience with this task. Nancy said when she gave her presentation, she was very nervous and did not do a good job (Nancy_EoSI). Tom did his presentation in week 15 together with two other boys. In the stimulated retrospective interview, he was not very happy with his own performance. He kept saying "my pronunciation is not very good"; when others could not understand him, he felt "very sad". As a student from a rural area, he had never had any training in listening or speaking before he came to university and his pronunciation was really poor. He even said he was "very disappointed" after hearing the teacher's feedback, because he felt that "the teacher said so much but I have no good points". It was true that Andrew talked a lot about pronunciation in his feedback, but before that, he praised Tom for having good "eye contact" and also "wearing a smile", but Tom did not seem to have noticed that, and instead felt "although the teacher didn't criticize me directly, there was hidden criticism". But he also admitted that he was angry with himself, not the teacher (Tom_SRI). Probably because of their comparatively lower language ability, this presentation task was a little too challenging for these two students and consequently affected their self-confidence negatively.

In contrast, Jack did not feel worried about his performance on this task because "this task didn't give me any pressure and I will pass this course whether my presentation performance is a little better or a little worse. But I want to do a good job because a good result will give me confidence and urge me to achieve more" (Jack_EoSI). When Jenny recalled her experience of giving her presentation with three other girls on the topic of 'hometown', she said proudly "when the teacher asked which presentation was the best, all the students said mine" (Jenny_EoSI).

Another point emerged regarding peer feedback after a student's presentation. After the week 15 lesson, when asked about his opinion on Andrew's feedback, Jack said he wished Andrew "had asked us to express our opinions", or rather Andrew could be more encouraging and supportive in helping students to express their ideas because he said "whenever I want to say something I become nervous, and then I am afraid of

speaking out my ideas" (Jack_SRI). In contrast, Jenny was hesitant about commenting on classmates' presentations. In the week 11 lesson, Jenny commented on one presenter's performance. She said "I think when he delivered the speech, he turned back to the screen"; however, she regretted having made such a comment to that classmate. "I am afraid I have hurt that student because I spoke out his shortcoming. … I think what I said was correct, but in our traditional Chinese culture, we'd rather say other people's good points than say their bad points. So I wonder if I have done something wrong. … I am really troubled" (Jenny_SRI). Clearly, while Jack wanted Andrew to scaffold more so that he could express his opinions more fluently and confidently, Jenny worried that her feedback might affect her relationship with her fellow students. However, it should be noted that both Jack and Jenny had fairly good language ability. According to classroom observation and interview data, the other volunteer students were not involved in peer feedback and they did not comment on this either. Therefore, more competent students were likely to be involved in peer feedback, but sometimes they were troubled that their comments might make their fellow classmates lose face.

4.2.3 The final exam

The profile of the final exam was generated from the analysis of Andrew's textbook, relevant documents (Appendixes 12, 13 & 14), and the teacher's and the students' end-of-semester interview data. This part follows the same pattern as that for the presentation task.

4.2.3.1 Assessment purpose

According to the minutes of the teacher group meeting (Appendix 13) held to prepare for the final test, the purpose behind the design of the final exam was to test what students had learned from the textbook, especially the "Inside View" section of the textbook, over the semester. "The design of the test topics should be based on the topics and the language functions covered in the Inside View section of each lesson. We should design one test topic for each lesson" (Appendix 13). Moreover, students' performance on the final exam accounted for 50% of the final grade (Appendix 12). Clearly, the final exam was designed for summative purpose.

In addition, the exam was conducted during the last week of the semester, after which Andrew did not have another class to meet the students again to provide them with some feedback on their performances at the test, and Andrew would not be teaching the same class the following semester either (Andrew_EoSI). Therefore, this exam served a summative purpose only.

4.2.3.2 Assessment method

The final test clearly fell into the category of *formal assessment*. Just as the instructions to the students (Appendix 15) specified, the interview data confirmed that the final exam taking the form of a semi-prepared role play was conducted at their usual class time in their own classroom during the last week of the semester. Students had been informed of all the seven role-play scenarios one week ahead, among which they were assessed on one by drawing lots (Andrew_COF14122010). In addition, during the test, Andrew followed the same procedure for each pair of students: after one pair of students drew their lot and was notified about their scenario, they waited outside and prepared for about five minutes during which another pair who had prepared for five minutes were assessed in the classroom, and this cycle repeated until all the students were assessed (Andrew_EoSI). During the testing time, Andrew also remained silent except for greeting and saying goodbye (Andrew_EoSI). Such kind of consistency could be regarded as a clear feature of formal assessment.

4.2.3.3 Making judgment

Who

The interview data showed that Andrew was the only one who made the judgment during the final exam. No other teachers or any students were involved in the assessment.

Construct

As specified in the minutes of the teacher group meeting (Appendix 14), the final exam was to test what students had learned from the textbook, especially the *Inside*

View section of the textbook, over the semester. Analysis of the final test topics and the textbook revealed a high level of agreement between the two in terms of format, content, and language functions.

Analysis of the *Inside View* section of each unit in the textbook revealed that after the warming-up and the listening-comprehension activities of two conversations, this section always ends with one or two role-play tasks which require students to make a conversation with a partner with the help of the clues provided in the tasks (A scanned copy of one unit of the textbook can be found in Appendix 16). This indicates that students are expected to learn to make up conversations, or to develop their *conversational competence*, a sub-competence of *interactional competence* in Celce-Murcia's framework (cf. Section 2.5).

In addition, this section of each unit always provides some cultural notes, some words and expressions of Everyday English, and some useful functional expressions, all of which are taken from the videoed conversations of that unit. Classroom observation showed that Andrew took the language functions of each unit as the learning objectives of that unit and would show them twice to the students, once around the beginning of a new unit, and then around the beginning of the second part of the unit as a kind of review (cf. Table 4.9). Andrew said he wanted to "draw students' attention to such language functions" (Andrew_SRI). Since students are expected to learn to perform certain language functions or speech acts, this *Inside View* section is intended to develop students' *actional competence*, a sub-competence of *interactional competence* in Celce-Murcia's framework.

Moreover, regarding *Cultural Notes* and *Everyday English* contained in this section, Classroom observation showed that Andrew always explained them to students. While Cultural Notes may help students understand the videoed conversations better, the words and expressions in the *Everyday English* box are essentially very common in spoken, contemporary English but may be difficult to understand (see Appendix 16 for an example). The fact that this *Everyday English* is a regular component of this Inside View section reflects the fact that this section is also intended to develop students'

formulaic competence, one component of Celce-Murcia's framework *of* communicative competence.

While the textbook, or the *Inside View* section of each unit in the textbook, provides opportunities for students to develop their interactional competence and formulaic competence, an examination of the test format and topics (Appendix 15) revealed that the final exam, in the form of a role-play, only required students to demonstrate their *conversational competence*, but the scenarios provided did not specify what language functions students should demonstrate or what everyday English expression students should use, thus giving students more room to demonstrate their *actional competence* and *formulaic competence*.

The final exam also had its marking criteria, which were the same as those for the presentation task (Appendixes 12 & 15). Therefore, as discussed before (see Table 4.6), the final exam assessed students' *linguistic competence*, *discourse competence*, *sociocultural competence*, and *cooperation skills*. Andrew said that he paid attention to the marking criteria when he graded students' performance during the final exam (Andrew_EoSI). He looked at "accuracy, variety and appropriateness of students' vocabulary and the variety in their language structures", which were about students' *linguistic competence*; he looked at whether students had "polite behaviours during their made-up conversations and good team work", which were about students' *sociocultural competence* and *cooperation skills*; he would also "try to look at if students can go deep into a topic and express their ideas with well-organized language so as to make their ideas or argument more convincing" (Andrew_EoSI), which was related to the depth of their ideas and their *discourse competence*. It can be seen that what Andrew claimed to be assessing had a higher level of consistency with the marking criteria than with what was emphasized in the textbook.

However, the fact that the test topics were given to students one week in advance to some extent caused the problems of construct underrepresentation and construct irrelevant variance (Messick, 1998).

Researchers have identified some basic characteristics of conversation, which include face-to-face interaction, unplannedness, potentially equal distribution of rights and

duties in talk, and manifestation of features of reactive and mutual contingency (Hughes, 2011; van Lier, 1989). However, the way Andrew's students prepared for the final test greatly reduced the spontaneity of their conversations. All the five interviewed students in their end-of-semester interviews said that they did "good" or "sufficient" preparation for the test. Jenny and Jack did not write out their dialogues word for word, but they did rehearse to make sure they would cover the key points and their dialogues would develop naturally. Tom said he and his partner rehearsed each of the seven dialogues for two or three times during the week before the test, "so at the exam time, what we said was more or less the same as what we rehearsed." Tina said she and her partner also prepared well for the test. "However, at the test time, he (Jenny's partner) asked one question in the wrong order. At first I was about to tell him that this was the wrong order, but then after a second thought, I decided to carry on. So I just put the answers to the two questions together." Nancy felt the final test was "quite easy," because she and her partner prepared well and during the test, "we two just said what we have rehearsed, and so that was good."

Actually, both Andrew and some students were aware of this problem. In the end-of-semester interview, Andrew specifically expressed his disagreement with the practice of giving students all the test topics one week beforehand, but he said he "was told to do so". However, he felt students should not get the topics in advance because

> I feel that (after they get the topics) they can prepare, and the way to prepare is to write out the dialogues and recite them. … As a result, they tended to over-prepare. You know the real purpose of oral communication is to express your ideas spontaneously in real-life contexts, so you won't have much time to think of very good sentences. Although during their preparation process, they might have tried to use as many expressions from the textbook as possible, and in that sense, this preparation process can be regarded as a review process, in the end, from the perspective of assessment, I still think we shouldn't give them the topics one week ahead. (Andrew_EoSI)

Therefore, Andrew doubted if such test scores could truly reflect students' oral English ability. Tina also felt that "because it allowed us to prepare in advance, the test was to some extent testing how well we had memorized our dialogues and couldn't test our ability to talk spontaneously" (Tina_EoSI).

Clearly, the fact that students prepared for the test and recited during the test had made the test unable to reveal students' spontaneous speaking skills. However, it should be noted that students were not just memorizing a script that had been written by someone else. When they prepared for the test, they did have to draw upon their linguistic competence, discourse competence, socio-cultural competence, and cooperation skills. Therefore, this memorization element had undermined the constructs intended to be assessed in this test, but not completely.

How

Regarding the scoring type, clear evidence showed that Andrew practiced analytical scoring. Having been given the marking criteria (Appendix 15), which consisted of three aspects—pronunciation and language quality, organization and content, and performance and teamwork—Andrew said he gave a separate mark on each of the three aspects for each student, because he felt in this way, "the marking can be more accurate" (Andrew_EoSI).

Regarding the basis of judgment, Andrew's decision-making process was essentially intuitive but also norm-referenced. Although he practiced analytical scoring, he still resorted to his own expertise to evaluate every aspect of a student's performance since there were no standards for him to refer to. Therefore, his judgment was intuitive in nature. In addition, when asked if he compared one student's performance against the other students', he said yes,

> When I mark a student, I pay attention that this is my impression of this student's performance at the moment. But I will also compare this student's performance with other students' performances. If student B has better oral English proficiency or better language quality than student A according to their daily performances, then for the final test, student A's score should not be too much higher than student B's. So I will make this kind of comparison, to try to keep balance. (Andrew_EoSI)

It can be seen that his general impression of his students still played a role in his decision-making, though he had tried his best to reduce its influence on him. Therefore, his marking was also NRM, though his decision was not only based on students' performances during the final test but also during their daily classroom performances.

There was no evidence to show that Andrew's scores were moderated by any external agents.

4.2.3.4 Making use of the judgment

Andrew reported that the test scores were used to generate the final composite scores for the students and he did not have another chance to provide any feedback to his students on their test performance (Andrew_EoSI). Therefore, his judgment was solely used for a reporting purpose.

4.2.3.5 Profile of the final test

The following figure summarizes the key features of the final test.
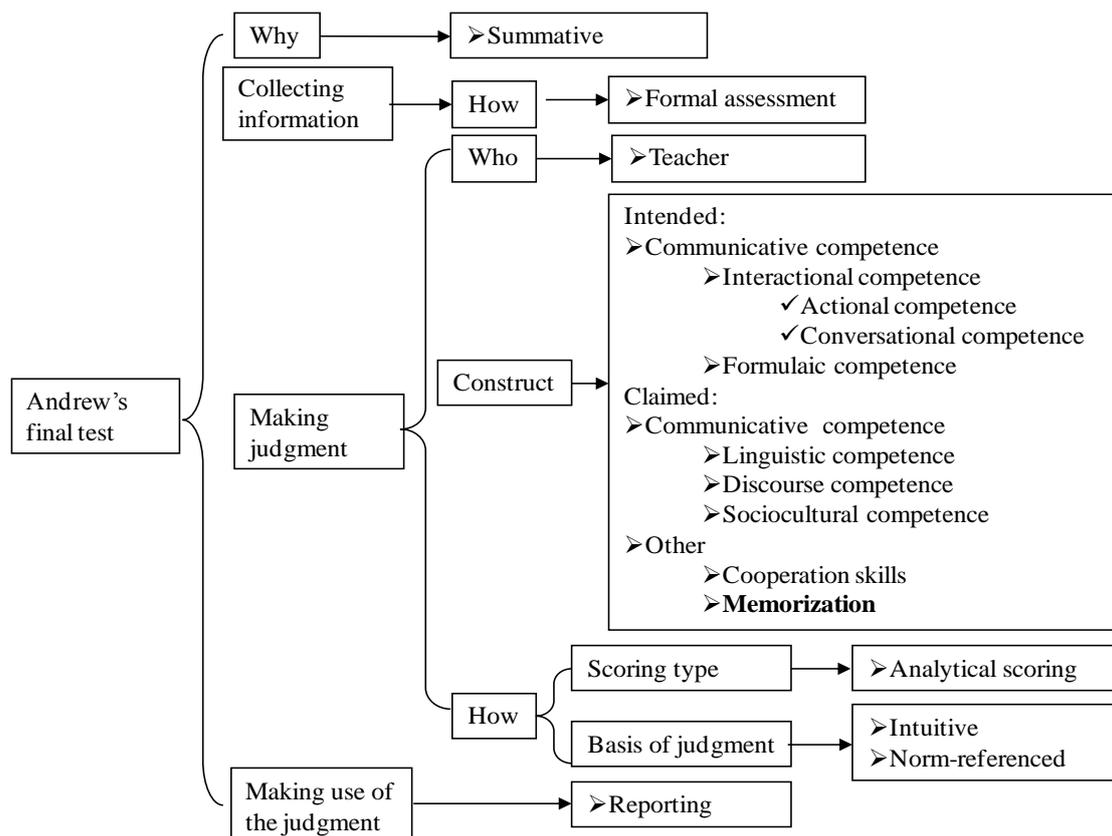
Figure 4.3 Profile of Andrew's Final Test

4.2.3.6 Students' perceptions

Although the test validity was found to be problematic, as discussed above, student interview data revealed that this final exam had some beneficial washback effects on the students and their learning.

First, the test itself served as a motivator for the five volunteer students to review what they had learned. Jack said "because we have to get the credit for this course, we are somewhat forced to practice. As a result, although the results may not be recorded in our transcripts, to some extent the test helped improve our oral English level. For me, I attach great importance to oral English, so I find it helpful, at least in increasing my desire to speak English" (Jack_EoSI). During the test, Jack "purposefully used some everyday English he had learned from the textbook" and he felt good about that (Jack_EoSI). Nancy liked the practice that students could get the test topics in advance because this "can make us practice more. You know there were many topics and this made us practice all of them. In the end, we were assessed on how well we had practiced those topics, and this is good" (Nancy_EoSI). Therefore, they took the preparation for the final test as a reviewing and practicing opportunity.

In addition, instead of causing anxiety in students, the test boosted some students' self-confidence in speaking English. Although the final test accounted for 50% in the final composite score, the test didn't cause much anxiety on the students' part. Before the test, both Tom and Jack in their stimulated retrospective interviews said they were not worried about the upcoming final test. Jack even added that

> It only accounts for 50%, and for the other 50%, I think I have got most of it. Like attendance and participation, I have done all of them. I don't think I will lose many points for those parts. And for this final exam, I think if I can give full play to my ability, though I may not achieve excellence, I should have no problem passing the test. (Jack_EoSI)

After the test, only Tina and Jenny mentioned that they were slightly nervous during the test while the other students were happy with their own performances. Nancy, whose oral English level was not very good, felt that the final test "enhanced my interest in learning English and made me more willing to talk in English, because after preparing for the final test, during the test I could speak complete sentences in English,

which really gave me a sense of satisfaction and made me feel happy and fulfilled"
(Nancy_EoSI).

The fact that this final exam stimulated students' learning motivation and enhanced
some students' self-confidence showed that even a test which is questionable in terms
of its validity can have a positive washback effect (Alderson & Wall, 1993).

4.2.4 Unrecognized CA

4.2.4.1 Data

A profile of Andrew's unrecognized CA practices was generated from the detailed
analysis of the classroom recording transcripts of the four observed lessons and the
relevant data from participants' interviews and student journals. Table 4.9 provides a
brief description of the classroom activities found when he taught the unit of *Sporting
Life*. (cf., Appendix 16)

Table 4.9 Andrew's Classroom Activities in Weeks 10 and 11

Duration in class	Classroom activity	Brief description
Week 10		
10:10-10:27	Student presentation*	Three female students gave their presentations and Andrew provided feedback
10:27-10:28	Introducing the learning objectives of this unit	Andrew showed a power point slide which contained the language functions that students were supposed to grasp
10:28-11:07: Warming up activities		
(10:28-10:42)	A vocabulary build-up activity*	Andrew elicited the names of different sports with the help of pictures shown on ppt. Some students volunteered some answers and then Andrew provided the correct answers.
(10:43-10:49)	Discussion*	Topic: Is your partner a sports fan? Format: Student pair work first followed by a teacher-led class discussion
(10:50-11:07)	Discussion	Topic: Suggestions for sports Source: *Starting Point* section of the unit on page 62 Format: Student pair work first followed by a teacher-led class discussion
11:07-11:15 Break		
11:16-12:00	*Inside View* section of the unit	1) Andrew explained the cultural notes, everyday English, and language functions contained in the textbook. Regarding conversation 1:

		2) Andrew drew students' attention to the map on page 63 of the textbook and asked students to give directions from Hertford College to Hertford College Boathouse. Two students attempted it but could not do the task well.
		3) Andrew asked the students to do exercise 3 on page 63 after watching the video of the first conversation.
		4) Andrew played the video of the first conversation.
		5) Andrew asked two students to give the directions again. The two students could not do this task well.
		6) Andrew played the video again.
		7) Andrew provided a version of giving the direction and then checked students' answers to exercise 3.
		8) Andrew played the video again, sentence by sentence, and explained the meaning of each sentence, and asked the students to repeat after each sentence. Regarding conversation 2:
		9) Andrew played the video of the second conversation.
		10) Andrew asked the class one question: What happened?* But there was no answer.
		11) Andrew played the video again.
		12) Andrew asked the same question again. Two students attempted some answers and Andrew provided feedback.
		13) Andrew played the video again, sentence by sentence, and explained the meaning of each sentence, and asked the students to repeat after each sentence. Regarding student practice/output:
		14) Andrew asked the students to do exercise 8 on page 64. (Students worked in pairs first. Then Andrew asked two pairs of students to perform before the whole class. After each pair's role play, Andrew provided feedback to the whole class.)
Week 11		
10:14-10:39	Student presentation*	Two male students gave their presentations and Andrew provided feedback
10:40-10:44	Reviewing	Andrew guided the students to review the language functions and the key expressions concerning giving directions that they had covered during the previous lesson by listing these expressions on two power point slides and going over them together with the students.
10:45-10:50	A vocabulary build-up activity*	Andrew showed a number of pictures of different sports and asked the students to provide the English names of these sports. Some students attempted some answers and Andrew provided the correct answers.
10:51-11:50	*Outside View* section of the unit	Regarding the videoed conversation: 15) Andrew asked students to do exercises 1 and 2 on page 66 individually and then led a class discussion. 16) Andrew asked students to do exercise 3 while watching the video and then played the video. 17) Andrew asked students what the conversation was about. There was no response from the students. Andrew explained the information about native speaker grammar on page 67. 18) Andrew explained the questions in exercises 3 and

		4 on page 67. Then he played the video again. Then he gave students some time to do exercises 3 & 4.
		19) Andrew played the video again, sentence by sentence, and explained the meaning of each sentence, and asked the students to repeat after each sentence.
		20) Andrew checked students answers to exercises 3 and 4 orally.
		Regarding student practice/output:
		21) Andrew asked the students to do exercise 5 on page 67. (Students worked in groups of four. Then Andrew asked two groups of students to perform before the whole class. After each group's discussion, Andrew provided feedback to the whole class.)
11:51-12:00	Discussion	Topic: Whether good sportsmen should be paid as much as film stars
		Source: One topic from the *Presentation Skills* section of this unit
		Format: Andrew asked students to discuss in groups of four. Then he conducted very brief class discussion.

A comparison between Andrew's textbook and the observation field notes revealed that except for a few activities, which are marked with a * in the above table, Andrew's classroom activities were almost completely based on the tasks and activities contained in the textbook. Andrew explained that he did so because he found the textbook very well designed.

> The textbook contained a rich variety of activities that were well sequenced from easier ones to more challenging ones and from giving students some input to asking them to output. … The earlier activities were to warm students up and prepare them for later activities. (Andrew_SRI).

Classroom observation also showed that the classroom activities of the week 14 lesson were very similar to those in the week 10 lesson, except that the week 14 lesson was about a new unit, *Job Fair*. It was understandable that the classroom activities of the two weeks' lessons were similar, because both lessons were dealing with the first half of one unit, each unit in the textbook contained similar tasks in a similar sequence, and Andrew's classroom activities were largely based on the tasks in the textbook.

Andrew explained that the way he conducted the week 11 lesson was his typical way of handling the second half of one unit (Andrew_EoSI). However, since week 15 was the last week of the semester, Andrew did not follow the traditional pattern. Instead, he devoted the most part of his class time to doing the *Unit Task* of that unit (Appendix 17), which he had assigned as a homework assignment the previous week

(Andrew_CO07122010). During the lesson of week 15, after reviewing the language functions covered in that unit, Andrew asked students to work in groups of four first before he asked three groups to present before the whole class and gave feedback after each.

In sum, the four observed lessons contained the following classroom activities (Table 4.10), excluding student presentation tasks and teacher explanations.

Table 4.10 Overview of Andrew's Classroom Activities

Unit	Week	Classroom activity	Percentage of class time within each lesson
Sporting Life	10	Vocabulary build-up activity	6.8%
		Two warming-up discussion activities	19.4%
		Listening comprehension checking	29.1%
		Role play	14.6%
	11	Vocabulary build-up activity	4.8%
		Warming-up discussion activity	8.7%
		Listening comprehension checking	34.6%
		Two oral discussion activities	19.2%
Job Fair	14	Vocabulary build-up activity	13%
		Two warming-up discussion activities	12%
		Listening comprehension checking	30%
		Role play	16%
	15	Role play	60%

It can be seen that over half of Andrew's class time was engaged in assessment-involving activities, though he did not regard them as assessment (Andrew_SRI). Besides, these activities fell into four types: vocabulary build-up activities, listening-comprehension checking activities, oral discussion, and role plays. In what follows, the assessment practices identified in the four types of activities will be described in detail.

4.2.4.2 Unrecognized CA practices

It should be pointed out that this group of activities showed common features in the following three aspects. First, Andrew's primary purpose was for pedagogical reasons rather than for assessment reasons, as can be seen in the following description of each type of activities. Since he had a strong desire to help students to learn and to improve and he usually provided timely and detailed feedback after students' performances in

class, the purpose of this group of CA activities was formative in nature. Second, Andrew had always been the agent of judgment, though occasionally, during some role-play activities, he would invite students to give peer comments. In general, students' involvement in the judgment-making step was very slight. Third, probably because Andrew did not consider these classroom activities as assessment, such categories in the descriptive framework (Figure 3.3) as *scoring type*, *basis of judgment, and moderation* were found irrelevant here.

Considering such commonalities and to avoid repetition, the description of this group of unrecognized CA practices will focus on the information-collecting method, the construct component of the making-judgment step, and the making-use-of-the-judgment step. Students' perceptions of each type of activity will be presented after the description of that type.

CA practices within vocabulary build-up activities

Classroom observation revealed that the typical way Andrew conducted the vocabulary build-up activities was through questioning and then providing feedback. This group of activities were all prepared in advance because Andrew always put the words he intended the students to learn on his ppt. While Andrew reported that his primary purpose of conducting such vocabulary activities was to "prepare students with some vocabulary that they might need in later discussion and to enlarge their vocabulary" (Andrew_SRI), the fact that Andrew provided feedback based on students' responses indicates that assessment was involved in such activities, as can be seen from the following excerpt.

Excerpt 4.1 is taken from his vocabulary activity during his week 10 lesson, during which he showed some symbols of the Olympic Games events on one power point slide and asked students to provide the English names of the sports events represented by the symbols. After he showed the slide, all the students got excited and tried to name those sport events in English. After a couple of minutes, Andrew began to go through the pictures one by one.

Excerpt 4.1

Turn	Speaker	Transcript	Feedback type
1	T	Now first let's have this test, ok. Do you know these sports?	
2	SS	Yeah, in Chinese.	
3	T	Yes, in Chinese, ok. As long as you can recognize them, that's good, you can use Chinese. So the first, what's the first one?	
4	SS	Skating.	
5	T	Skating, that's good. ((pointing to the next symbol))	Confirmative
		[...]	
18	SS	Dancing	
19	T	Trampoline 蹦床 ((beng chuang, trampoline)), not dancing.	Corrective
…		[...] ((going through all the 35 symbols))	
116	T	… The last one is modern :: ?	
117	SS	((no response))	
118	T	Pentathlon, ok. 我们先看这个词啊 ((wo men xian kan zhe ge ci a, Let's look at this word first)), Pentagon 是什么 ((pentagon shi shen me, What does pentagon mean))?	Corrective
119	SS	((no response))	
120		Pentagon 五角大楼对不对？((It means pentagon, right?)) So, pentan 就是什么？((pentan jiu shi shen me, What does pentan mean?))	Corrective
121	SS	五 ((wu, five))	
122	T	Five, ok. Pentathlon就是什么呢((pentathlon jiu shi shen me, What does pentathlon mean?))...	Corrective
123	SS	现代五项 ((xian dai wu xiang, modern pentathlon))	
124	T	现代五项 ((xian dai wu xiang, modern pentathlon)), modern pentathlon. Ok.	Confirmative

In this excerpt, Andrew elicited students' vocabulary knowledge of the major events in Olympic Games by asking questions (turns 1 and 3). When students could provide the correct words (turn 4), Andrew usually just confirmed their answers (turn 5) and moved on to the next one. When students did not know the correct word (turn 18), Andrew often provided the correct answer and sometimes added the Chinese translation (turn 19). When students did not have a clue to a word or expression, as with "modern pentathlon" in the above excerpt (turn 117), Andrew not only provided the correct answers but also provided some clues to help students understand and memorize the word/expression, as shown in turns 118 and 120. It can be seen that Andrew's feedback varied depending on students' actual responses rather than just telling students the answers regardless of students' responses. Therefore, assessment was involved in such activities. The fact that these activities were all prepared in

advance made such vocabulary activities fall into the category of *planned assessment-involving instructional activities*.

Further analysis revealed that assessment actually occurred when Andrew provided feedback to students' responses to each word. Specifically, around each word/expression, there was a sequence of teacher initiation, student response, and teacher feedback, though sometimes the first step was not verbalized but was made through Andrew's body language, like his pointing at a picture rather than asking a question. Such episodes were usually very short and essentially took the form of the traditional IRF (Initiation-Response-Feedback) sequence (Sinclair & Coulthard, 1975). Considering that Andrew prepared those words beforehand, such assessment episodes around each word should be regarded as *planned assessment-involving instructional episodes*.

Obviously, such vocabulary activities focused on students' vocabulary knowledge, a kind of linguistic competence. Andrew's feedback reflected the fact that Andrew used his judgment to promote student learning by providing corrective (14 instances), confirmative (26 instances), or explanatory (11 instances) feedback.

The volunteer students in general loved this type of activity. For example, Jack wrote in his journal that the most impressive moment of the week 10 lesson was "when Andrew conducted the Q&A session on names of different sports, which I heard of but hadn't known how to say them in English. Now I have memorized such words as judo, equestrian very clearly" (Jack_J1). Tina said the vocabulary activity helped her "to recall some of the words she had learned before" and she felt that "in future when I have to talk with someone about sports, then it won't be that I know nothing of it. … So I find this activity very necessary" (Tina_SRI). Nancy and Jenny also found the vocabulary activities "useful" and they had copied down some of the words in order to look them up after class (Nancy_SRI; Jenny_SRI).

CA practices within listening-comprehension checking activities

Each unit in Andrew's textbook contains two conversations in the *Inside View* section and one conversation in the *Outside View* section, and around these conversations are

121

a number of comprehension checking activities, which take the form of multiple choice items, answering questions, and filling in blanks (cf. Appendix 16). Classroom observation showed that Andrew conducted almost all of the comprehension checking activities and his usual way was through assigning students tasks to do and then providing feedback after students finished their tasks. Although Andrew's primary purpose of conducting such comprehension-checking activities was to "help students understand the material" rather than assessing their listening comprehension (Andrew_SRI), there was evidence to show that the detailed-ness of Andrew's feedback varied according to students' responses, which indicates that assessment was involved in such activities, as can be seen from the following excerpt.

This excerpt is taken from one of the listening-comprehension checking activities during his week 11 lesson. He had asked students to do exercise 3 while they watched the video. After he had explained the videoed conversation sentence by sentence and students had repeated the conversation sentence by sentence, Andrew checked their answers to exercise 3 (see Appendix 16).

Excerpt 4.2

Turn	Speaker	Transcript	Feedback type
1	T	Do you get all the answers to the questions?	
2	SS	((Some students nodded their heads.))	
3	T	Ok, uh, first one, Tae Kwon Do originated about::?	
4	SS	2000	
5	T	2000 years ago in Korea. Number 2,Tae Kwon Do is basically the::	Confirmative
6	SS	Hand and foot	
7	T	Hand and foot martial arts. Number 3, it's open to::	Confirmative
8	SS	((no responses))	
9	T	Anyone. We take any flexibility level, any age, and any:: fitness level. Ok, if you are very fit, 很健康对不对((hen jian kang dui bu dui , very fit, right)), if you are kind of overweight, or if you are not flexible, you are not so healthy, you could also go, ok, so any level are welcome. Number 4.	Corrective & explanatory

In this excerpt, Andrew elicited students' comprehension of the listening material by asking questions (turns 1 and 3). When students could provide the correct answers (turns 4 and 6), Andrew usually just confirmed their answers (turns 5 and 7) and

moved on to the next one. When students did not know the correct answer (turn 8), Andrew provided the correct answer and added some explanation and Chinese translation (turn 9). Therefore, it can be seen that Andrew's feedback was contingent upon students' responses, which indicates that Andrew evaluated students' responses and adjusted his feedback accordingly. The fact that all such listening-comprehension activities were taken from the textbook made such activities *planned assessment-involving instructional activities.*

Further analysis revealed an embedded-ness of assessment episodes centering around each question within each listening-comprehension activity, similar to the pattern found in those vocabulary activities. Specifically, when Andrew provided feedback to students' responses to each question, there was an IRF sequence. Considering that all the comprehension questions were from the textbook, each IRF sequence should be regarded as a *planned assessment-involving instructional episodes*.

Obviously, this group of activities assessed students' listening comprehension, part of students' *linguistic competence*. During the activities, Andrew assessed how well students' could cope with the comprehension questions and provided *task-referenced* feedback to help with student learning. Analysis showed that for this group of activities, there were altogether 23 instances of confirmative feedback, 3 instances of corrective feedback, and 4 instances of explanatory feedback.

Some volunteer students found such comprehension-checking activities useful in helping them "grasp the main idea of the conversation" (Nancy_SRI), and "improve the note-taking skills" (Tina_SRI), and Andrew's feedback helped them "check whether my comprehension was correct" (Jenny_SRI).

CA practices within discussion activities

Classroom observation showed that Andrew conducted discussion activities both as warming-up activities around the beginning of a lesson to activate students' background knowledge about the topic and as extension activities around the end of a lesson to allow students to use what they had learned to express their true opinions (Andrew_SRI) (cf. Table 4.9). His usual way of conducting such discussion activities

was through assigning a topic or a task to students. After students had practiced in pairs or groups for some time, he would conduct a teacher-led discussion, during which he would first pose the discussion topic or question again, and after one or some students gave their responses, he often provided his feedback. While Andrew did not consider such activities as assessment activities, analysis showed that he often provided feedback after students' attempted answers. This showed that assessment was involved in such activities, as can be seen from the following two excerpts.

Excerpt 4.3 is taken from the warming-up discussion of the week 10 lesson, and the topic was "Is your partner a sports fan?". Andrew organized students into groups of four and asked them to practice in groups first before he asked some students to report to the whole class about their group members. Excerpt 4.3 shows the first student's report to the class about the sports habits of his group members.

Excerpt 4.3

Turn	Speaker	Transcript	Feedback type
1	T	Ok, so have you found whether your partner is a big sport fan or not? Have you gone through all these questions? Now I'm gonna ask some of you to report to the whole class what you know about your partner in terms of sports? Does she or he like sports or what kind of exercise habits does she or he have? ((Andrew moved towards one group in the front of the classroom and talked to one boy in the group.))	
2	T	Ok, so tell us your partners, ok. Does she or he like sports? What do you find out about your partner? Ok.	
3	S1	This girl? ((indicating the girl sitting beside him))	
4	T	Yeah.	
5	S1	This girl never do exercise and::	
6	T	She never <u>does</u> exercise?	Corrective
7	S1	Yeah yeah yeah. And this girl ((indicating the girl sitting behind him)) almost once a week just like me. And this girl	
…		[...] ((Andrew asked the boy about the names of his group members and required the boy to use their names instead of "this girl". The boy checked with the three girls in his group. The next turn is about what the boy said concerning the third girl in his group.))	
17	S1	She like to play badminton, she do she play once a week and she is not a very big sports fan.	
18	T	She is not a big sports fan, and she <u>plays</u> badminton once a week?	Corrective
19	S1	Yeah.	

| 20 | T | That's not bad. ((Andrew turned towards another students.)) Ok, what about you? What did you find out about your partner? | |

It can be seen that in excerpt 4.3 there were two instances of corrective feedback (turns 6 & 18). Specifically, when S1 said something that contained a grammatical error (turn 5), Andrew reformulated the sentence and provided a correct form (turn 6). In turn 18, Andrew repeated what S1 had just said, but in a slightly more precise way, since S1's version was a little wordy with three grammatical errors. Both examples showed the feature of recast, because Andrew did not state explicitly that an error has been committed (Sheen & Ellis, 2011). However, obviously, this kind of feedback indicated that Andrew evaluated S1's performance and his judgment was used by providing a recast to the student. Therefore, Andrew was engaged in assessment though he did not recognize it as such.

Excerpt 4.4 is taken from exercise 5 in week 11 lesson. This exercise was for students to express their own opinions, after they had studied the video material in which a Tae Kwon Do coach was interviewed to talk about the history of this sport and the advantages of practicing it. This exercise asked students to express their opinions on four questions relating to sports in general. It was designed to develop students' critical thinking ability, as specified in the textbook (cf. Appendix 16). This excerpt is about the class discussion on the first question: what sports can offer people, which is not exactly the same as the first question listed in the textbook.

Excerpt 4.4

Turn	Speaker	Transcript	Feedback type
1	T	Ok, I have to stop you guys. Have you finished? Ok, let's see the first one, so 'apart from physical exercise, what can sports offer people'? What can sports offer people? ((indicating one boy sitting in the first row to talk))	
2	S1	We can make more friendship and we can be more confident.	
3	T	Uhha, confidence and friendship. What else do you guys think?	Confirmative
4	S2	To build up your self-confidence	
5	T	To build up the self-confidence, ok, good. ((fixing his gaze to S3))	Confirmative
6	S3	Also helps develop business	
7	T	Uhha ((He seemed to be waiting for more information from	

125

8	S3	Of course, uh: it helps us relax and make good friends.	
9	T	Yeah, good, that's a very good point because after playing tennis, like after a week's work, I feel much stressed. So after playing tennis, though I sweated a lot, it's fun. I feel like I am totally relaxed after I play tennis. Ok. ((fixing his gaze to S4))	Confirmative & explanatory
10	S4	I think you can be more energetic to do your job or study.	
11	T	That's right. Ok. Exercise helps you to build up your energy capacity, like your capacity is this small, but after exercise, your capacity is this big, you know ((he made a gesture to show the difference)). So even after your work or study, you use much of your energy, but you still have some energy left. So that's very important. Ok. Let's look at the next one.	Confirmative & explanatory

(the student.)) appears at the top of the first partial row above row 8.

It can be seen from the above excerpt that Andrew focused on the ideas of what students said, and he sometimes confirmed (turns 3 & 5) and sometimes even elaborated on their ideas (turns 9 & 11). The nature of his feedback reflected that he assessed whether students' responses were relevant and then he used his judgment to provide feedback accordingly. Therefore, it should be proper to say that such discussion activities involved assessment although Andrew regarded such activities as "a natural extension of the content we've just covered, and at the same time, students could practice some informal discussion skills I have introduced to them" (Andrew_SRI). Therefore, such discussion activities are regarded as *planned assessment-involving instructional activities*.

Further analysis also revealed many embedded assessment-involving episodes during the feedback stage of such discussion activities after the students had discussed the assigned questions in groups or pairs. Unlike the vocabulary activities and the comprehension-checking activities where those assessment-involving episodes were centred around each word or each question, for the discussion activities, the assessment episodes were centred around individual students. Sometimes, for the same question, Andrew would ask several students to express their opinions and provided feedback respectively. Considering that all the discussion questions were prepared before the class since Andrew often listed the discussion topics/questions on his power point slides, such assessment-involving episodes should be regarded as *planned assessment-involving instructional episodes*. Moreover, occasionally, Andrew also

provided online comments to help students improve their language accuracy (e.g., turn 6 in Excerpt 4.3), which was an *incidental assessment episode*.

From the nature of his feedback on such discussion activities (cf. Excerpts 4.3 and 4.4), it can be seen that Andrew focused more on students' ideas than on their language quality, though occasionally he did give feedback on students' language accuracy. Analysis revealed that there were 7 instances of corrective feedback where he corrected students' language errors and 10 instances of confirmative feedback and 11 instances of explanatory feedback where he provided feedback on students' ideas. His judgment obviously served the purpose of promoting student learning.

Some students found such discussion activities useful because they were "opportunities for us to practice our oral English and present ourselves before others" (Tina_SRI), and they could "enrich my ideas and stimulate me to think" (Jenny_SRI). Jenny even added that while she was doing such discussion activities, she had to "pay attention to her pronunciation, to her language accuracy, and to her ideas", which she found "very good, very challenging" (Jenny_SRI). Moreover, some students felt nervous when Andrew was around observing and some did not feel so. Jack said "when the teacher is observing me I will feel nervous, worrying that I may not speak well or speak loud enough" (Jack_J1). Tina said in her stimulated retrospective interview that she was a little nervous when the teacher was around observing because she was afraid that she might not speak well. In contrast, Jenny did not feel nervous when the teacher was observing her. Instead, she wished Andrew would give her more attention (Jenny_J1). She wanted Andrew to find out her problems from what she had just said.

CA practices within role-play activities

Andrew sometimes conducted role-play activities, usually after the careful study of the videoed conversations in the *Inside View* section. Such activities were all from the textbook and were intended to be opportunities for students to practice the language functions and useful expressions they had just learned (cf. Table 4.9). While Andrew did not regard such activities as assessment (Andrew_EoSI), analysis of classroom

recording and observation field notes showed that assessment was involved in such activities, as can be seen from the following excerpt.

Excerpt 4.5 is taken from exercise 7 in the week 14 lesson. This is the last activity in the *Inside View* section after a careful study of the video material. In this task, students were asked to make up a dialogue talking about their plans for future jobs. Detailed instructions about how to do this dialogue were provided in the textbook (Appendix 18). Andrew gave students seven minutes to work in pairs. Then he asked two pairs to perform before the class. Excerpt 4.5 is about the second pair's performance and Andrew's feedback.

Excerpt 4.5

Turn	Speaker	Transcript
1	T	And then, you two, yes, two gentlemen. ((indicating the two boys to come to the front of the class))
2	S1	What do you want to be when you finish your college?
3	S2	Well, I just don't like did a job about language.
4	S1	So what do you be?
5	S2	Maybe a lawyer or a manager.
6	S1	But you don't study laws in the college.
7	S2	I will sure to study economy or math in the second year of my university life.
8	S1	In here?
9	S2	Yeah. Actually I can't study in other universities, you know. So what about you?
10	S1	I want to go to some other countries after I finish this school and studying.
11	S2	So what countries do you like?
12	S1	USA or German, or Belgium.
13	S2	So you are live there or just come back?
14	S1	I just come back. China is good now.
15	S2	To do what?
16	S1	To do something I don't know.
17	S2	I think you'll have a promising future.
18	S1	Ok, thank you. ((Other students applaud.))
19	T	Ok. Good. **An interesting sincere discussion**, but yeah, I mean in real life, you talk, you face each other, but because this is also a kind of presentation, you have an audience. That's why **you have to tilt your bodies a little bit**, to face them, not exactly facing them, but just tilt your body a little bit. Ok, uh also as I said before, when you do presentation, like when you do this kind of dialogue, conversations, **try to use the expressions you just picked, from the conversations we watched**. Otherwise, you are only using what you've learned before. You are not really practicing what you just learned. Do you get my point? Ok.

In the above excerpt, turn 1 performed the function of eliciting student performance, although Andrew had elicited students' performances by asking students to make up and practice their dialogues in pairs first. After the two students' performance, Andrew's feedback (turn 19), which was explanatory in nature, showed that he had evaluated the two students' dialogue, as can be seen from the bolded parts, and based on his judgment he provided some feedback, not only to the two students, but also to the whole class. Because the task was from the textbook and Andrew's feedback contained clear signs of his judgment of the evidence he collected and his use of the judgment he made, this episode was regarded as a *planned assessment-involving instructional episode*

Further analysis revealed that assessment often occurred during the feedback stage of such role-play activities after the students had practiced in groups or pairs. Due to the big class size, each time Andrew only asked two or three pairs of students to perform before the class. Considering that all the role-play activities were from the textbook and Andrew often asked some students to perform in front of the class before he provided his feedback, the role-play activities should be regarded as *planned assessment-involving instructional activities.*

Analysis of Andrew's feedback on students' role-play performances revealed that Andrew looked at students' sociocultural competence, paralinguistic competence within interactional competence, linguistic competence, formulaic competence, as well as the content of students' dialogues (see Table 4.11). His feedback was always task-referenced and mainly explanatory (11 instances), though he did provide corrective feedback twice. Therefore, such CA practices were used to promote students learning.

Table 4.11 Aspects Mentioned during Andrew's Feedback on Role-plays

Aspect	Example
Sociocultural competence (4)	But in a formal interview, you need first start with a very polite greeting, ok, good morning, a handshake, and then you ask questions.
Paralinguistic competence within interactional competence (3)	When you give an interview, you will face the situation, you have to go for an intervi: w, uh, so the ways, you manner, your gesture, you know, your voice, they all say something about you, so you have to be careful about that.

Formulaic competence (3)	Good! Uh, the two girls, they've used some expressions to:: to sympathize, to congratulate. That's very good.
Linguistic competence (2)	You could say "what do you like to do after your graduation" instead of "graduating".
Depth of idea (1)	And the interviewees they gave very good answers to very good questions, I mean I have to say the interview was quite tough, the questions were quite tough, but they gave very good answers.

Regarding students' perceptions, students generally agreed that it was a good practice opportunity, but sometimes one's partner might influence how much one could practice. Nancy felt it a useful way to practice oral English, because it was not very difficult and she could use many expressions she had just learned (Nancy_SRI). Tina felt that, "at university, you always need a partner to learn English, whether you like it or not. So cooperation is very important", and this role play activities helped improve her ability to cooperate with other students (Tina_EoSI). Jack also said that the unit task, a group role play about job interviews, was to practice students' cooperation skills, but because he and his group members were from different departments and they seldom saw each other after class, he couldn't prepare for the assignment before class though he wished to (Jack_SRI). As for the activity itself, he liked it but found it hard to think of the content and sometimes he didn't have the language to express what he wanted to say (Jack_SRI). Tom liked the group interview task (Unit Task for the unit of *Job Fair*) very much in the first place, but working with his group members made him feel very uncomfortable.

> They had very good pronunciation. They talked and talked. I couldn't understand and I couldn't get in. … I and another student, our English is not as good as theirs. So they two just kept talking and we two didn't talk much. You know the teacher said four students worked in a group. So in fact I liked this activity very much, but I didn't do it very well. (Jack_SRI).

4.2.4.3 Profile of Andrew's unrecognized CA practices

The following figure summarizes the key features of Andrew's unrecognized CA practices. The numbers after specific types of feedback show the frequencies of respective types of feedback found in the data.
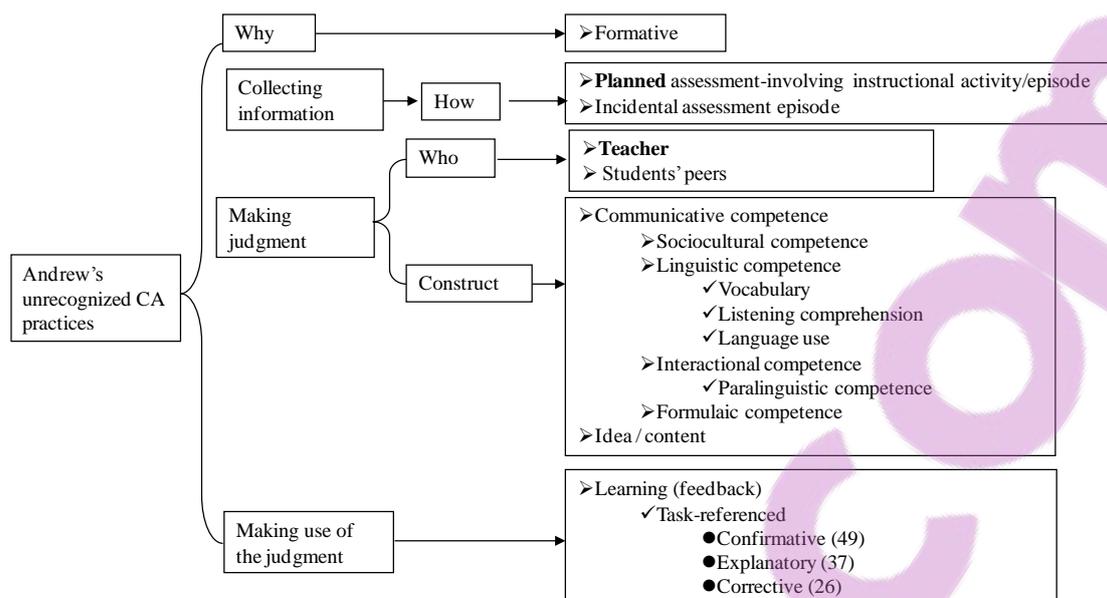
Figure 4.4 Profile of Andrew's Unrecognized CA Practices

4.3 Students' general perceptions

Independent sample t-tests on part II of the beginning and the end-of-semester questionnaire data showed that over the semester, students' mastery goals increased significantly and their performance avoidance goals decreased significantly, while the other four aspects showed no statistically significant changes (Table 4.12).

Table 4.12 Andrew's Students' Self-reported Oral English Learning over the Semester

Aspect	At the beginning of the semester		At the end of the semester		Mean difference
	No. of students	Mean	No. of students	Mean	
Self-evaluation	29	19.97	26	20.08	+0.11
Mastery goals	28	12.61	27	13.85	+1.24**
Performance approach goals	28	11.46	26	12.53	+1.07
Performance avoidance goals	29	6.59	27	5.07	-1.52*
Anxiety	29	8.07	27	8.22	+0.15
Effort	28	11.29	27	11.19	-0.1

Note: *: $p<0.05$, **: $p<0.01$

Analysis of the data from part III of the end-of-the-semester questionnaire showed that students' perceptions concerning Andrew's classroom practices were quite different from the findings obtained from classroom observation and relevant interview data. In the following table, the numbers in the "frequency" column are the numbers of students who made a specific choice. The last column is the author's judgment of Andrew's classroom practices based on the four observed lessons and the relevant interview data. When there was no evidence to determine the frequency of a particular practice, the author put in "unclear".

Table 4.13 Students' Perceptions of Andrew's CA Practices

| Please circle the practices that your teacher has used in this oral English course. | Frequency | | | | | Information from classroom observation and interview |
	Once the whole semester	Every month	Every unit	Every lesson	Every activity	
1) Explain the learning objectives of the whole course.	3	8	6	2	7	Once only
2) Explain the learning objectives of a particular unit.	3	2	19	0	0	Every lesson
3) Explain the learning objectives of a particular activity.	12	2	5	0	1	Every lesson
4) Explain the connections between the learning objectives of the whole course, of a particular unit, and of a particular activity.	2	6	3	4	2	Unclear
5) Explain the skills to be grasped in this course.	5	10	8	0	0	Unclear
6) Organize activities for students to practice the skills to be grasped in this course.	2	15	7	0	1	Every lesson
7) Observe how well students have grasped the skills.	6	10	7	2	0	Every lesson
8) Ask questions to check how well students have learned.	5	16	4	1	0	Every lesson
9) Comment on students' performances after they finish a classroom activity.	12	11	3	0	0	Every activity
10) Point out students' strengths and weaknesses after they finish a classroom activity.	14	11	1	1	0	Every activity
11) Tell students the correct answers after students finish a classroom activity.	9	12	3	1	0	Every lesson
12) Tell students how to do better next time after students finish a classroom activity.	6	10	4	3	0	Every lesson
13) Praise those students who perform well in doing classroom activities.	8	11	2	1	1	Every lesson
14) Criticize those students who do not try their best to do a classroom activity.	0	2	2	2	8	unclear
15) Encourage students not to be afraid of making mistakes.	7	8	6	2	0	Unclear
16) Organize students to do self-assessment.	1	12	2	2	3	Unclear
17) Organize students to do peer-assessment.	3	10	1	1	4	During some activities but not in every unit
18) Assign after-class homework.	2	15	7	2	0	Every lesson
19) Check students' homework in class.	1	4	5	5	2	Occasionally but not in every unit
20) Explain to the students how they are assessed in this course.	1	1	1	1	17	Once only
21) Inform students about what they are expected to do at the course test(s).	0	3	2	0	13	Once only
22) Explain to students how they can get high scores in the course test(s).	0	0	3	0	14	Unclear

133

As can be seen from the above table, except for the six items about which no conclusive evidence was available, inconsistency was prevalent between students' perceptions and the author's judgment based on the available data. One possible reason for such discrepancy might be that not every student understood the questionnaire items in the same way or in the way the author understood them, which undermines the validity of the questionnaire data. In light of this, considering that students' perceptions of the usefulness of Andrew's CA practices were based on their understanding of Andrew's CA practices, no further analysis was conducted to the part of the data concerning the usefulness of the selected practices. Instead, this questionnaire was further revised and then used in Linda's class.

4.4. Summary

In this oral English course, Andrew was constantly engaged in a variety of assessment practices throughout the semester. However, he only regarded those formally conducted assessment activities that led to final grading as assessment, namely the final test and the presentation task, but would not regard those FA practices embedded in his daily classroom teaching (cf. 4.2.4) as assessment.

Consequently, there was a clear line drawn between FA and SA in his CA practices. While Andrew could feel that he formed general impressions of his students' oral English abilities from their classroom performances, he intentionally guarded against their influence on his judgment when he was doing summative grading. Therefore, FA and SA practices were as clearly separated as possible in this course.

Such kind of separation between FA and SA can be explained by Andrew's beliefs on assessment and learning.

> I don't want to use tests to make my students work harder. Actually I have always been telling them that you should make the learning process an enjoyable process. Don't focus on the tests. Otherwise, you will become too goal-oriented and you will miss the fun of the learning process. (Andrew_EoSI)

Therefore, although the presentation task was originally designed for the washback purpose, he never told his students their scores even though he did score them and keep a record.

Regarding the formal assessment tasks, though they were given to him by his teaching group rather than designed by himself, he exerted some agency when conducting the presentation task. On the one hand, he turned this task into a learning opportunity by allowing students sufficient time to perform and providing them with detailed and constructive feedback. Thus, students found this task very beneficial for their learning. On the other hand, to both comply with the assessment requirements from his teaching group and to stick to his own belief that assessment should be fair and objective, he used the grades for students' presentations to represent their class participation scores.

However, teaching in a "high structure" context (Wette & Barkhuizen, 2009), he sometimes felt he had to do what he was asked to do, even if he had to compromise his own belief. This can be seen from his attitude towards the final test. Although he believed that the final test topics should not be given to students one week in advance because students might be "over-prepared" and the test then might not truly reflect their language proficiency, he still did what he was told to do.

In addition to the two formal assessment tasks, Andrew was also engaged in a large number of CA practices during his everyday teaching, which remained unrecognized by him and his students. His unrecognized CA practices were embedded within the feedback stage of four types of his classroom activities: vocabulary build-up, comprehension checking, discussion, and role plays. This group of CA practices were mostly planned assessment-involving instructional activities and episodes, though there were occasional incidental assessment episodes as well. His planned assessment-involving instructional episodes and incidental assessment episodes were mainly in the form of IRF sequence.

In all his CA practices, Andrew had always been the major agent of judgment. He occasionally invited the students to make comments after some students' presentations, but in most cases, students were not involved in peer-assessment or self-assessment.

About types of scoring, which was relevant only to the formal assessment, Andrew practiced analytical scoring for the final test but holistic scoring for the presentation task. About the basis of judgment, since no specific standards were provided concerning what scores should be given for what kind of performances, Andrew had to

resort to his own expertise to evaluate students' performances. Therefore, his marking was essentially intuitive. However, for the final test, in order to make sure his marks were fair, he also practiced NRM. For his classroom-embedded CA practices, NRM or CRM was not that relevant at this micro level, and his ongoing judgment was more expertise-based.

Finally, the analysis revealed that Andrew assessed more than just students' communicative competence. Through both FA and SA practices, Andrew not only looked at many aspects of students' communicative competence (Celce-Murcia, 2008), especially their sociocultural competence, linguistic competence, interactive competence, discourse competence, but also assessed students' cooperation skills, use of visual/aural aids, vocabulary knowledge, listening comprehension, and depth of students' ideas.

Chapter 5. Mary's CA Practices

This chapter describes the results of the data analysis for Mary's case. It follows the same structure as that for Andrew's case. After the description of her teaching context, Mary's formal assessment structure is presented first. Then the profiles of the components within her formal assessment structure are presented separately before the profile of her unrecognized CA practices are described. Students' general impressions of her CA practices as revealed from the students' questionnaire data are presented after that. This chapter ends with a summary of the typical features of her CA practices.

5.1 Mary's teaching context

5.1.1 Educational background and past teaching experiences

Mary, a native speaker of English from the US, had a Masters degree with some training in TESOL. She did her Master's degree in her late thirties in the field of Spanish language with a minor in linguistics in order to become an ESL/EFL (English as a second/foreign language) teacher. After she got her MA, she taught for a few years "international students who came to the US to study as university students and also immigrants who were trying to settle down in America" (Mary_BI). During those years, she went to some TESOL conferences and attended some TESOL workshops. She moved to China in 1999 and stayed in Beijing until the time of the present study. By then, she had worked at four different universities in Beijing where she mainly taught language skill courses such as reading, writing, and speaking, but sometimes also some content courses such as American Geography (Mary_BI). The other four universities are all universities with some kind of specialty such as forestry, sports, medicine, and law, and they are all much less prestigious than this university, which is a top university in China and a comprehensive university. At the time of the study, it was her second round of teaching in this university.

5.1.2 Her university, her department, and her students

The present study was conducted at the English Department of the university where Mary was teaching. As a leading centre for teaching and research in China, this university embraces "diverse branches of learning such as basic and applied sciences, social sciences and the humanities, and sciences of medicine, management, and education" (from the university website, accessed on Jan. 20, 2012). Together with the other 15 departments of the School of Foreign Languages of this university, the English Department "commits itself to providing high-quality citizens and experts for China and the world, those who can adeptly fit into globalized working environments and serve the interest of the human society with outstanding foreign and native language competence and thorough understanding of foreign and Chinese cultures" (from the university website, accessed on Jan. 20, 2012).

First-year English-major students in this department had four compulsory language-skill courses in their first semester: Intensive Reading (4 hours/week), Intensive Listening (2 hours/week), Extensive Listening (2 hours/week), and Oral English (2 hours/week). In addition, students could choose a variety of optional courses, both in Chinese and in English, both from the English Department and from other departments or other schools. The English Department offered many optional courses in English to first-year students such as Greek and Roman Mythology, Structure of English, Bible Stories, etc.

This English Department receives about 50 students each year, divided into three parallel classes. Mary was teaching oral English to all the six classes from year one and year two. The class I observed was one first-year class. There were 18 students, five male and 13 female. None of them came from foreign language high schools but they all had very high scores in English in their College Entrance Examination before they entered this university. Three students started to learn English before the age of 8, two since becoming middle school students (aged 13-15), and the remaining since primary school (aged 9-12). One student had lived in America for one year before she started her primary school, four other students had travelled abroad for less than a month, and the rest had no experience of travelling abroad.

Data from the second part of the beginning questionnaire showed that most students in this class were moderately confident about their oral English, as can be seen from items 1 to 4 in Table 5.1, but they were not satisfied with their present oral English ability, as can be seen from items 5 to 7.

Table 5.1 Mary's Students' Self-Evaluation of their Oral English Ability at the Beginning of the Semester

Questionnaire item	Not true at all	Not really true	Partly true	Mostly true	Absolutely true	Mean
1. I am very confident when I speak English.	1 (5.6%)	4 (22.2%)	7 (38.9%)	6 (33.3%)	0	3
2. I worry that native English speakers will find my oral English strange.	4 (22.2%)	9 (50%)	1 (5.6%)	2 (11.1%)	2 (11.1%)	2.39
3. My oral English is very fluent.	1 (5.6%)	5 (27.8%)	7 (38.9%)	5 (27.8%)	0	2.89
4. My poor oral English always makes me feel inferior to those who can speak fluent oral English.	3 (16.7%)	6 (33.3%)	3 (16.7%)	5 (27.8%)	1 (5.6%)	2.72
5. When I speak in English, I can always find the words to express my ideas.	2 (11.1%)	9 (50%)	5 (27.8%)	2 (11.1%)	0	2.39
6. When I speak in English, I can say exactly what I want to say.	1 (5.6%)	8 (44.4%)	6 (33.3%)	3 (16.7%)	0	2.61
7. I am satisfied with my present oral English proficiency.	8 (44.4%)	7 (38.9%)	2 (11.1%)	1 (5.6%)	0	2.17

Similar to Andrew's students, at the beginning of the semester, Mary's students had high mastery goals and performance approach goals but low performance avoidance goals (Table 5.2).

Table 5.2 Mary's Students' Goal Orientations at the Beginning of the Semester

Questionnaire item	Not true at all	Not really true	Partly true	Mostly true	Absolutely true	Mean
Mastery goals						
8. It's very important for me to completely grasp the skills taught in this course.	1 (5.6%)	0	2 (11.1%)	2 (11.1%)	13 (72.2%)	4.44
12. I am willing to spend a lot of time practicing to improve	0	0	2 (11.1%)	4 (22.2%)	12 (66.7%)	4.56

my oral English.						
15. I believe that if I work hard, my oral English will improve.	0	0	0	6 (33.3%)	12 (66.7%)	4.67
Performance approach goals						
9. It is very important for me to be the top student in this oral English class.	0	2 (11.1%)	3 (16.7%)	7 (38.9%)	5 (27.8%)	3.88
11. I am eager to become a smart student in my teacher's eye in this course.	0	2 (11.1%)	4 (22.2%)	3 (16.7%)	9 (50%)	4.56
14. I wish to get a high score in this course.	0	0	2 (11.1%)	4 (22.2%)	12 (66.7%)	4.56
Performance avoidance goals						
10. I will be satisfied so long as I can pass this course.	8 (44.4%)	9 (50%)	1 (5.6%)	0	0	1.61
13. I worry that I may fail this course.	7 (38.9%)	6 (33.3%)	4 (22.2%)	0	1 (5.6%)	2
16. I feel that no matter how hard I try, my oral English will make little improvement.	14 (77.8%)	4 (22.2%)	0	0	0	1.22

The beginning questionnaire also revealed that students' anxiety levels varied but not very high (Table 5.3) and most students put much effort to improve their oral English, especially in improving their grammatical accuracy and enlarging their oral English vocabulary (Table 5.4).

Table 5.3 Mary's Students' Anxiety Level at the Beginning of the Semester

Questionnaire item	Not true at all	Not really true	Partly true	Mostly true	Absolutely true	Mean
17. I get very nervous when a foreigner asks me something in English.	0	3 (16.7%)	4 (22.2%)	6 (33.3%)	5 (27.8%)	3.72
18. I feel distressed about being unable to improve my oral English.	8 (44.4%)	3 (16.7%)	3 (16.7%)	4 (22.2%)	0	2.17
19. I worry that other students will laugh at me when I speak English.	8 (44.4%)	6 (33.3%)	0	4 (22.2%)	0	2

Table 5.4 Mary's Students' Effort at the Beginning of the Semester

Questionnaire item	Not true at all	Not really true	Partly true	Mostly true	Absolutely true	Mean
20. I actively create opportunities to talk with others in English.	0	7 (38.9%)	5 (27.8%)	3 (16.7%)	3 (16.7%)	3.11

21. I regularly enlarge my oral English vocabulary.	0	1 (5.6%)	7 (38.9%)	7 (38.9%)	3 (16.7%)	3.67
22. I try my best to improve the grammatical accuracy of my oral English.	0	0	6 (33.3%)	6 (33.3%)	6 (33.3%)	4
23. I grasp every opportunity to practice my oral English.		1 (5.6%)	10 (55.6%)	4 (22.2%)	2 (11.1%)	3.41

5.1.3 The oral English course

In this department, the tradition was that the Oral English course should be taught by native speakers. When the study was conducted, Mary was teaching this course for the second round. Mary had been given total freedom to decide what to teach and how to teach because the department did not set specific requirements on what students should achieve and just told her "This is the class and this is the time. Go there and teach" (Mary_BI).

Therefore, Mary had no textbook for this course.

> When I applied, the Department told me they didn't have a textbook. If I want to use a textbook, I have to find it myself. I have to pick something they have in the bookstore here because it's easy to get for the students. And it should be cheap since the students don't want to pay for it. So last year, the first year, I suggested a textbook and they all complained that it was too expensive. They wanted to have a photocopy, but that's illegal. (Mary_BI)

About when to teach what in this 16-week course, Mary had a syllabus but only in her mind (Mary_BI). Based on classroom observation and interview data, she did not ask students to do role plays but mainly gave students topics for discussion. Over the semester, they talked about family, education, internet dating, business, Halloween and ghost stories, Thanksgiving, Christmas holiday and *Christmas Carol* (the cartoon film based on Charles Dickens's novel). Mary provided all the topics and students were not invited to bring in topics that they were interested in, though she planned to do it the next time she taught this course (Mary_EoSI).

About the course objectives, Mary, after some pause, decided that this course was to enhance students' fluency in speaking English, because she noticed that "many first-year students had never ever spoken out loud in English in front of other people, especially with a native speaker" and she should "facilitate opportunities for the

students to practice English" in this course, to make them "use English in class as much as possible, to improve their fluency" (Mary_BI). At the end of the semester when she reflected on the course, she rephrased the objective of this course as to get the students "comfortable speaking in English, loosen them up somewhat, and get them comfortable talking and having conversations and speaking about things that were beyond what they've already spoken about" (Mary_EoSI). In her mind, "accuracy was not the major goal. I think of course it's needful, but I think for most of them they just need to open their mouths and speak" (Mary_EoSI).

It can be seen that Mary's understanding of fluency was mainly concerned with students' confidence in speaking, which is different from what fluency is usually understood in the language teaching and language assessment field, where fluency is generally used to mean a naturalness of flow of speech, reflected through "amount of speech, rate of speech, unfilled pauses, filled pauses, length of fluent runs between pauses, repairs, clusters of disfluencies" (Wood, 2010, p. 17). Moreover, she thought her course objective was to provide opportunities for students to practice so as to enhance their confidence in speaking in public, not in terms of specific language skills and/or standard of performance.

About student assessment, in the middle of the semester, Mary spent five weeks on two types of speeches, which served as the midterm test for the students. The first week was a kind of preparation, and during the next two weeks, each student gave an "impersonated speech", that is, each student read aloud a self-selected famous speech in a way similar to the original speaker, and during the next two weeks, each student gave a speech of their own on the topic: the most meaningful thing to me. The last week of the semester was used for the final test (Mary_EoSI).

It can be seen that, while Mary claimed that she focused on "fluency", the "impersonated speech" task was clearly designed to assess and hopefully enhance student accuracy in pronunciation. In addition, classroom observation revealed that during classroom teaching, Mary provided a lot of feedback relating to the accuracy of pronunciation, grammar, and vocabulary use. This showed that Mary also focused on the accuracy aspect of students' oral performance.

Overall, Mary had a lot of freedom to decide on what to teach, when to teach, how to teach, and how to assess, except that Mary kept mentioning one policy from her university, that is "only 40% of the students in a class can have a score over 85" (Mary_SRI, Mary_EoSI). To a great extent, Mary was working in a "low structure" environment because "the curriculum pre-specifications were minimal and flexible, allowing teachers and learners to negotiate the curriculum" (Wette & Barkhuizen, 2009, p. 198).

For students, this Oral English course was a required course, and students' performances in this course would affect their comprehensive evaluation at the end of the semester, which in turn would affect if a student could win a scholarship.

5.2 Mary's CA practices

5.2.1 Formal assessment structure

As the only teacher for this course and without any instruction or requirements from her department except for the university assessment policy mentioned above, Mary worked out her own assessment structure from her past teaching experiences. "I usually give 50% for the final test, 30% for the midterm speech, and maybe 20% for some other activities", because "we should not base the whole mark on the final exam" (Mary_BI).

Classroom observation and interview data showed that Mary certainly regarded the midterm test and the final test as assessment and derived the final scores mainly from students' performances on these two tests, but had an ambivalent attitude towards taking students' performances on classroom activities as the basis for assessment. On the one hand, during these classroom activities, she was "usually too involved in what the conversation is" to keep track of students' performances (Mary_EoSI). But on the other hand, Mary acknowledged that she did form a general impression of a student according to their daily classroom performances and she generally put a student into one of three big categories: "above-average", "just-average", or "below-average" (Mary_EoSI).

In addition, Mary's interview data showed that she did not calculate a final composite score according to the percentages mentioned above. Instead, she only "compare[d] how they did in the final with how they did in the class, and tr[ied] to bring out an average" (Mary_EoSI). Here "how they did in the class" mainly referred to students' performances on the midterm test, about which she kept careful records, but not students' performances on daily classroom activities during which she was too busy interacting with the students to grade them (Mary_EoSI). For the final composite scores, while she mainly averaged students' midterm and the final test scores, she also made slight modifications according to her general impressions of her students based on their classroom performances, because she felt she was obliged to do so. "I have to do that [make modifications], cos as I said already we can only have 40% above 85. So if I have too many people above 85, then I have to look at them and decide well who do I need to bump down" (Mary_EoSI). Furthermore, classroom observation also showed that Mary checked students' attendance sometimes by calling out student name list and noted down those who were late or absent for a class. She said students' absence or being late for class would also affect their final scores (Mary_SRI). The following figure summarizes Mary's assessment structure in her oral English class.
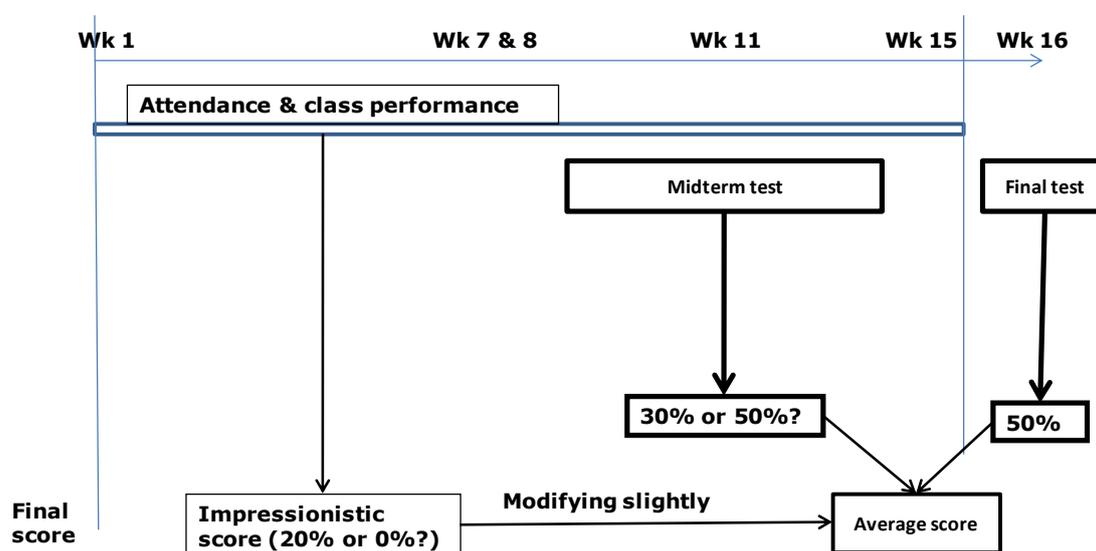
Figure 5.1 The Assessment Structure in Mary's Oral English Course

While Mary did not regard her classroom activities as assessment, analysis revealed that assessment was involved in some of her classroom activities and some episodes of

her interactions with her students. In what follows, the profiles of the midterm test and the final test will be presented before the description of those unrecognized CA practices.

5.2.2 The midterm test

The profile of the midterm test was generated from the analysis of classroom recording transcripts and classroom observation field notes relating to the midterm test, relevant interview data, and student journal data. Using the framework outlined in chapter 3 (Figure 3.3), this midterm test is described from the following four aspects: assessment purpose, information-collecting method, judgment-making process, and making use of the judgment. Students' perceptions about this task are described after the profile of this task.

5.2.2.1 Assessment purpose

The midterm test consisted of two parts: students reading aloud a self-selected famous speech in a way similar to the original speaker, or "impersonating a speech" in Mary's term, and students giving a prepared speech on a given topic: The most meaningful thing to me.

Interview data showed that Mary's primary purpose in conducting the two tasks was for grading. Mary said that during her teaching, she was "very involved" with her teaching and found it "difficult to watch them [the students] in class and to analyze their progress in class" (Mary_EoSI). Therefore, she "had them do those two speeches" as a way "to single out the students and look at them objectively" (Mary_EoSI). Moreover, Mary believed that she was "just deciding where they are and their level of English" instead of "what they learned in my class" (Mary_EoSI). Therefore, the midterm test helped Mary make decisions about students' language proficiency and mainly served *summative* purposes.

In addition, data showed that the impersonated-speech task was actually an add-on because Mary did not have this part in her original plan. She did plan to ask students to give a speech, as was her usual practice, but as the course proceeded, she soon noticed

that pronunciation was a problem for most of her students, so she decided to add this task to "help them with their pronunciation" (Mary_EoSI).

> I think just fluency is the goal. And then I noticed pronunciation was a big problem, so I kind of I noticed in some cases pronunciation was an obstacle to them being understood. And so that became a focus. (Mary_EoSI)

After the midterm test, Mary gave students a handout on the typical pronunciation problems that Chinese students tend to have (Daniel_EoSI, Lucy_EoSI). This showed that this impersonated-speech task was included in the midterm test for the purpose of help with student learning, and therefore also served *formative* purposes.

5.2.2.2 Assessment methods

Classroom observation and interview data showed that for both speeches, students prepared in advance and in class they took turns to stand up in the front of the classroom and read their speeches or gave their own speeches individually, followed by Mary's feedback. The impersonated-speech task was essentially a reading-aloud activity of a self-selected speech. For students' own speeches, classroom observation showed that all the students who gave a speech that day talked to the whole class instead of reading from written scripts, sometimes with the help of an object, such as a postcard or a photo album, or with power point slides. It can be seen that the midterm test was a *formal assessment* (Mary_COF23112010).

5.2.2.3 Judgment-making process

Who

In the week 11 lesson, three students gave their impersonated speeches and six students made their own speeches before the class. Classroom observation and recording showed that only Mary provided feedback after each speech and there was no peer evaluation or self-evaluation involved. At about ten minutes before the class should be over, when all the students who were to give their speeches that day had finished, Mary started a class discussion, asking students to comment on the six speeches they had heard, but none of the students did so, and this discussion turned out

to be about students' general feelings about giving a speech, their worries and strategies to cope with their worries (Mary_COF23112010). It can be seen that Mary had the intention to invite students to give peer comments, but probably either because this discussion time was not so close to the specific speeches or because Mary's feedback had been given and students found it hard to come up with some new comments, this part of discussion was not related to the six speeches. Therefore, this part could be regarded as a general feedback section after students' performances rather than a peer-evaluation session.

Construct

For the impersonated-speech task, Mary intended it to be an opportunity for students to practice their pronunciation (Mary_EoSI), and analysis of her feedback on the three impersonated speeches in the week 11 lesson showed her feedback did focus on students' pronunciation skills, such as liaison, stress, pause, etc. (see Table 5.5). Obviously, this impersonated-speech task assessed students' pronunciation skills, one aspect of their linguistic competence.

Table 5.5 Aspects Mentioned during Mary's Feedback on Students' Impersonated Speeches

Aspect	Example
Individual sounds (4)	Buddies, my friends, right? Buddies. Bodies is you know what that is anyway. /o/ /ʌ/ ((she makes the two sounds distinct)) bodies, buddies, two different things, so it was confusing. ... Ok, so body buddy body buddy. I love my buddies over there. Take care of my body. ((laugh)) Take care of my buddy. Take care of my body.
Liaison & stress (4)	You tend to have that choppy Chinese style of pronouncing a word, ok, rather than having them connect, and you know, stressing some and de-stressing others, you tend to give all the words the same stress.
Pause (2)	I thought you did very well with the pauses, and stressing, and yeah, whenever you use pauses, as I've mentioned already before, it shows a lot of drama in the speech, you know, it helps to impress us.
Accent (1)	You really got the sense of American English, so that was good!

For the prepared-speech task, Mary intended to use it to assess students' speech-making techniques.

> Before I had them give their speeches, I gave them a handout on good speaking techniques. There are like six different techniques, and they totally ignored that, the whole thing. Yeah, it's like when they came to give their speech, maybe the

time distance was too far in between, or maybe I had to reinforce it, because they totally ignored it. (Mary_EoSI)

Unfortunately, the researcher could not get a copy of the handout on good speaking techniques. Therefore, the analysis was mainly based on the actual feedback Mary provided after the six speeches in the week 11 lesson. The analysis showed that Mary mainly focused on students' speech content, speech organization, pronunciation, and language style, which reflected that Mary assessed three components of students' communicative competence: discourse competence, linguistic competence, and sociocultural competence, but she also assessed students' use of visual aids and their ability to meet the time limit (see Table 5.6).

Table 5.6 Aspects Mentioned during Mary's Feedback on Students' Own Prepared Speeches

Aspect	Example
Content (7)	
Content relevance (4)	So that was a really interesting topic, and that could be a very effective speech if you worked on it a little more because you know we can all relate to that, right?
Content significance (3)	<u>You have some very good content</u> here, so really meaningful message. Your message is very meaningful.
Discourse competence (4)	
Idea Progression(4)	I like the way you started and connected it to a meaningful experience in life. Good organization.
Linguistic competence (4)	
Pronunciation (4)	I'd like just to mention that your pronunciation is good. I don't know if you've been practicing that, but you have really good connection.
Sociocultural competence (1)	
Stylistic appropriateness (1)	I would comment, and probably American students need to hear this too, that there are different, we use different kinds of language in different settings, right? So there is really informal language that we use with our friends, and there is more formal language you would use when you are giving a talk or a presentation, and but it's hard to break some habits, sometimes we have a habit that we use informal English that don't fits so well in a more formal atmosphere like when you are giving a speech. Yeah, and those would be like adding like uh, a very terrible habit American teenagers have, well not just teenagers, anyway, but yeah, so those are the kind of thing you want to avoid when you are giving a presentation, yeah.
Other aspects (2)	
Use of visual aids (1)	I like your picture. Your picture helps. Anytime you use some kind of visual aids, like the letters, the picture, yeah, yeah, exactly understandable. It's effective. It helps us pay attention as well. Yeah, so that was good.
Time limit (1)	Part of speech making is you sticking within a time limit. Yeah, that's part of making a speech. Sometimes you have to choose what you are going to speak about and what you are not going to speak about, right? Yeah, so you have to choose ahead of time.

How

Mary had no marking criteria for either speech. Though she gave feedback on many aspects of students' performances during the two speeches (cf. Tables 5.6 & 5.7), she said she usually just gave one score instead of separate scores for different aspects (Mary_EoSI). Therefore, she practiced holistic scoring.

Regarding the basis of her judgment, Mary mentioned that to meet the university assessment requirement, she had to compare her students against each other and rank them because "we have to put them in a hierarchy" (Mary_EoSI). This showed that her assessment was *norm-referenced*. In addition, Mary said her marking was usually "intuitive" (Mary_EoSI), which implies that her decision-making was expertise-based.

There was no evidence that her marks on students' speeches were moderated in any way.

5.2.2.4 Making use of the judgment

As mentioned in section 5.2.1, the midterm test was one of the two major sources for the final grading of students. Therefore, it was used for the reporting purpose.

In addition, analysis showed that the midterm test was also used for learning purposes. Specifically, after each speech, Mary often provided feedback (cf. Tables 5.6 & 5.7), sometimes very detailed, to help with student learning. Her feedback was all task-referenced, and explanatory (26 instances altogether) and corrective (4 instances) in nature.

Moreover, she provided students with a handout on the typical pronunciation problems based on students' performances during the midterm test. The fact that she did not prepare this handout before the test but after the test showed that probably she realized that pronunciation was still a serious problem for some students and she should give them something extra for them to refer to after class so that they could practice their pronunciation by themselves. It can be seen that her judgment was also used formatively to modify her teaching.

5.2.2.5 Profile of the midterm test

Figure 5.2 shows the key features of this assessment practice.
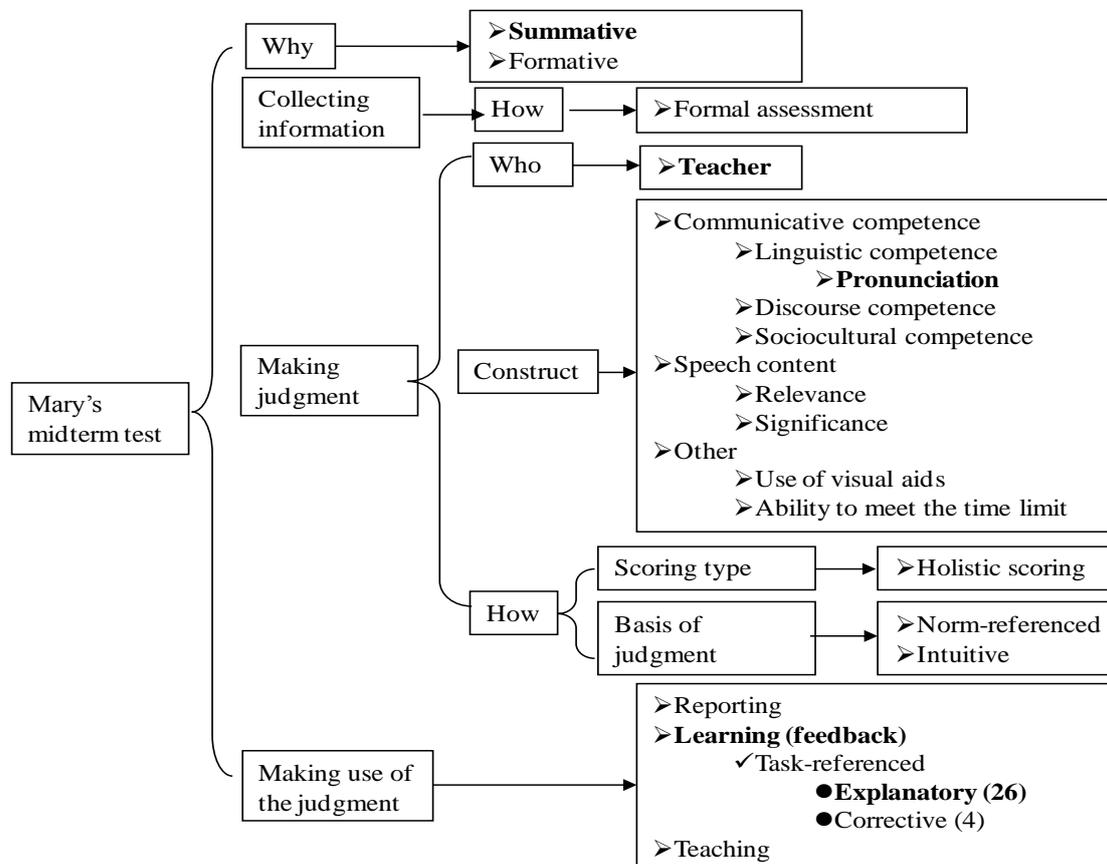
Figure 5.2 Profile of Mary's Midterm Test

5.2.2.6 Students' perceptions

Four students volunteered to participate in the journal writing and student interviews and their information can be found in the following table.

Table 5.7 Background Information on the Volunteer Students in Mary's Class

Student name	Gender	English-learning starting age	High school type	Going-abroad experience	Oral English ability
Daniel	Male	13-15	Non-FL	No	Average
Vivian	Female	Before 8	Non-FL	1Y in the US	High
Lucy	Female	9-12	Non-FL	No	Low
Jody	Female	Before 8	Non-FL	1 month tour in Australia	Average

Student journals and interview data showed that students all learned from the midterm test, but at the same time also learned for the test.

For example, Vivian gave a speech on being a volunteer for Beijing Forum. She had a lot of power point slides but was cut short because she exceeded the time limit. After her speech, Mary pointed out that her language style was too informal and might be inappropriate for a speech. In Vivian's journal, she wrote,

> From Mary's feedback, I have learned that I should learn more of formal language and every sentence I say should be meaningful and be worthy of its being said. Also, I should put across the most important thing first and after that I can talk about details. (Vivian_J3)

Daniel did not give a speech in the week 11 class and was mainly a listener. He wrote in his journal that "I have learned some strategies in giving a speech, such as pausing, stressing, putting your own emotion into your speech, and having confidence. Giving a speech is actually a process of sharing your ideas with others and you should enjoy it" (Daniel_J3). He also talked about his understanding of the differences between speech and presentation because he found some students' speeches were actually presentations.

> Giving a speech requires speech delivering skills, which should mainly be language-related. Your language should be coherent and powerful. A presentation is mainly for the purpose of introducing or presenting something, and the main methods to do it are through language, sounds, videos, and it doesn't have a high requirement on your language being coherent and powerful. (Daniel_J3)

This showed that he did learn from other students' performances. Besides, when talking about his own preparation for this task, he decided that "when I give my speech, I will pay attention to my word connection, pronunciation, stress, etc., but of course, the most important thing is the main idea" (Daniel_J3). Here, he was taking in Mary's feedback because Mary emphasized students' pronunciation and main idea when she gave feedback on students' speeches.

The idea of recording one's own voice and listening to it was the suggestion from Mary after Lucy's impersonated speech (Mary_COF23112010). Obviously, students

did learn from their own performances, other students' performances, and the teacher's feedback.

Lucy, who gave an impersonated speech and whose oral English was comparatively poor, focused more on the score she got than on Mary's feedback, but the score also impressed on her that she had pronunciation problems. In the stimulated retrospective interview, she said she was satisfied with her performance, but after she secretly found out her score, she became clearly aware of her own pronunciation problem.

> I think I did a satisfactory job today. … I took a secret look at the score Mary gave me. I only got an 8. I saw some student got 8.5, and some got 9. … I cannot say this 8 is a disappointing score for me, … but it makes me realize that I am lacking something. So I think I still have a very very big problem with my pronunciation. … I saw Mary noted down 'difficult to understand', … so I know my pronunciation is not very accurate. (Lucy_SRI)

Probably that score had such an impact on her that she even did not pay attention to what Mary said after her speech. Her comment on Mary's feedback was "I think what the teacher said was just to make you feel confident, such as 'excellent' and things like that, but she didn't score according to what she said" (Lucy_SRI). It was true that immediately after Lucy's speech, Mary said 'Wonderful! Good!', but then she gave long feedback, pointing out one pronunciation error, then suggesting that the students should record their own reading and examine it to improve their pronunciation, and finally pointing out another problem: adding syllables. Finally Mary said 'good' again (Mary_CRF23112010). Obviously this score had a greater effect on Lucy than Mary's feedback.

In fact, not only Lucy but all the other volunteer students, despite their different language proficiency levels, cared a lot about the test scores, because this test score would affect both their final scores and their GPA, which was very important for students at this university, as explained by Vivian and Lucy.

> GPA is everything at university. It will determine whether you will be guaranteed to a master's program without taking the required exams, whether you are allowed to do a second degree, whether you can pursue further study abroad, and whether you can find an ideal job in future. (Vivian_J3)

You know I have always been thinking how to raise my GPA. ... To be frank with you, you know there are about 40 students in our department, and when we reach year 3, only five students will get the chance to study abroad as an exchange student for half a year. I have been thinking how I can become an exchange student then. But according to my present scores, there is no hope at all. I am almost the worst student in the class. (Lucy_SRI)

Therefore, since the test scores meant a lot for the students, the test pushed them to be test-oriented to some extent.

5.2.3 The final test

Since no documents were available concerning the final test, the profile of the final exam was generated from the analysis of the recording transcripts, classroom observation field notes, and the teacher's and the students' end-of-semester interview data. This part follows the same pattern as that for the midterm test.

5.2.3.1 Assessment purpose

The final test was not designed in advance but evolved naturally from Mary's teaching. Students' end-of-semester interviews showed that students were not informed of the final test until one week before the end of the semester. In the middle of the week 15 lesson, Mary informed her students that, instead of asking them to discuss all the topics covered during the semester, she would ask them to talk about one of the following two topics during the final test: their reactions to the film *Christmas Carol* they just watched but had not discussed in detail yet, and the meaning of their impersonated speeches, since they had not had an opportunity to talk about those speeches in class (cf. Appendix 19).

Such a decision was based on Mary's understanding of the purpose of the final test and her intention to meet the university requirement.

Yeah, tests are kind of necessary evil, because for the university system, we have to categorize them, we have to focus on what they've accomplished, we have to put them in a hierarchy, it's too bad, actually. But that's the nature of our society. But yeah, so, I think the final somewhat measure their accomplishment, their ability. (Mary_EoSI)

It can be seen that Mary regarded the final test as a way to summarize students' achievement in this course, thus summative in nature.

5.2.3.2 Assessment method

In the middle of the week 15 lesson, Mary informed her students about the form of the final test. Students would choose their own partners and during the test students would come in pairs and have a discussion for about ten minutes on one of the two topics assigned by Mary: reactions to the film *Christmas Carol* or the meanings of their impersonated speeches. The test would take the form of a paired discussion but sometimes Mary would join their discussion and make it a "three-way conversation" (see Appendix 19).

Classroom observation of the final test confirmed that the test mainly took the form of a paired discussion. During the discussion, Mary mainly sat there observing and taking notes, but sometimes she would join their discussions by asking some questions based on what the students had just said (Mary_COF28122010). After each pair finished their test, Mary usually said some nice words about the students' performances, such as "You guys do great!" even if the students had not performed that well, and then had a little chat with them about their arrangement for the upcoming winter holiday before the students left. It can be seen that the final test was a formal assessment.

5.2.3.3 Making judgment

Who

Classroom observation showed that Mary was the only examiner, and the students were not involved in any peer assessment or self-assessment (Mary_COF28122010).

Construct

There were no specific marking criteria for the final test. Analysis of how Mary graded the first two pairs of students (Appendix 20) revealed that she considered three aspects before she reached a score for a student: the student's linguistic competence, the depth

of his/her ideas, and her general impression of the student's previous classroom performances.

Regarding linguistic competence, Mary considered: 1) fluency, as when she said "Clair, she is a little choppy but she was able to keep going"; 2) pronunciation, as when she said "I would guess Cathy I would give her 80. She did well but she had a lot of problems, a few problems with the pronunciation and stuff"; and 3) grammar, as she noticed Clair "had some grammar problems and pronunciation problems" so that she "couldn't quite understand her" (Mary_CRT28122010).

In addition, Mary sometimes checked students' listening comprehension.

> Sometimes I like to interact with them cos I can tell whether they understand me or not. I mean I noticed the couple, the two girls that didn't understand something I had said to them. So both Cynthia and Liya they had some comprehension problems. (Mary_CRT28122010)

Sometimes she asked the students to start and maintain their discussion by themselves because she wanted to see their "interlocution skills" (Mary_CRT28122010). This aspect was about students' *conversational competence*, one component within the interactional competence in Celce-Murcia's model (2008).

Besides, Mary considered the depth of the content of what a student said. For example, she said "Andy's content was much more in-depth than Cathy's," one of the factors that led to her decision that she would give Andy "an over 85" but Cathy "80". For Clair, besides her accuracy and fluency problems, the fact that "the depth (of her ideas) wasn't that great" also contributed to Mary's decision that she should get a score "below 80, or 78 to 80" (Mary_CRT28122010).

Mary's general impression of a student's ability also played a role in her decision-making. For example, when she was grading Andy, she said "Andy, he is very smart, a little shy and a little quiet, but that's not a problem"; when she was grading Vivian, she said "Vivian's English of course is very good. Yeah, she's been in America for a while, so she got a high score" (Mary_CRT28122010). It can be seen that to some extent, Mary tried to make her scores match her general impression of her students based on their daily classroom performances.

155

Moreover, the fact that the test topics were related to what they had discussed in class before and were given to students beforehand might have entailed the risk of students memorizing for the test. However, Mary thought this was a fair way to do the final test because she thought the test should "measure how much students have achieved" over the semester and "it wasn't quite fair if I would just have them talk about something we hadn't talked about, or we were not assessing what they had practiced in class", so she tried to "cover things we have covered" (Mary_EoSI). To minimize the memorizing problem, Mary emphasized beforehand that those who recited during the test would get a low score (Mary_COF21122010) and during the test Mary occasionally joined the students' discussion to make the discussion more spontaneous rather than pre-planned (Mary_COF28122010). Also, Mary only told the students the general topics for discussion and the specific discussion topic and questions were determined by Mary at the test time; thus the students had no time to prepare but had to start their discussion immediately. Such practices might also enhance the spontaneity of the discussion. The end-of-semester interviews with the volunteer students showed that while the students all prepared for the test, what they actually said during the test was different from what they had prepared, because they only "prepared some ideas that should be relevant" to the discussion instead of the "exact words to be said" (Jody_EoSI).

In brief, in the final test, Mary mainly examined students' linguistic competence and the depth of their ideas, but her judgment was influenced by her impression of a student's overall English proficiency she got over the semester. In addition, the fact that students could prepare for the test might have reduced the cognitive and linguistic demand involved in the discussion task, thus causing the problem of construct underrepresentation (Messick, 1998), but not to such a great extent as that of Andrew's final test.

How

During the test time, there was a ten-minute break after the first pair of students and another ten-minute break after the second pair, because no students had signed up for those time slots. Therefore, during these two periods, Mary and I had a chat about how

she graded the first two pairs of students (Appendix 20). Her words showed that she adopted holistic scoring and her basis of judgment was norm-referenced.

> I try to grade them, I mean, I have fluency, pronunciation, grammar, vocabulary, and content, and I have tried, because I have to give them a percentage score, so I have tried to do it mathematically, but it never works out in the end. When you add them all up, it's like either too high or too low or too many have the same, so I dispensed with that.

So when she graded a student's performance,

> I am trying to think of their fluency, their pronunciation, grammar, if they have a lot of grammar mistakes, I'll put some negative marks, but you know it's only intuitive in the end, just intuitive. It's really hard to calculate, you know, you just lose track of how much you need to take off to get them in the right range. So sometimes you take off too much and sometimes too little, so it's just hard. It's hard.

It can be seen that Mary found analytical scoring troublesome because it was ineffective in ranking the students correctly. Her concern for putting students in the right order was evidenced again when she talked about how she graded the first pair of students.

> I would guess Cathy, I would give her 80. She did well but she had a lot of problems, a few problems with the pronunciation and stuff. Andrew I think did very well, yeah, though a little bit shy. Andrew, he is very smart, a little shy and a little quiet, but that's not a problem. So I probably give him an over 85. Yes, I am <u>just not sure. I need to compare with the other students...I thought Andrew's content was much more in-depth than Cathy's</u>. (Emphasis added)

Here Mary was not only comparing the two students against each other, but also was prepared to compare them with the other students. Clearly, her basis of judgment was norm-referenced. Although she considered many aspects in her evaluation, she did not evaluate students' performances against any standard, and therefore, her basis of judgment was not criterion-referenced.

5.2.3.4 Making use of the judgment

Student interview data showed that students were not informed of their final test scores. Instead, they only received the final composite scores early the next semester.

Therefore, Mary's judgment on students' performances during the final test was used solely for the reporting purpose.

5.2.3.5 Profile of the final test

The following figure summarizes the key features of the final test.
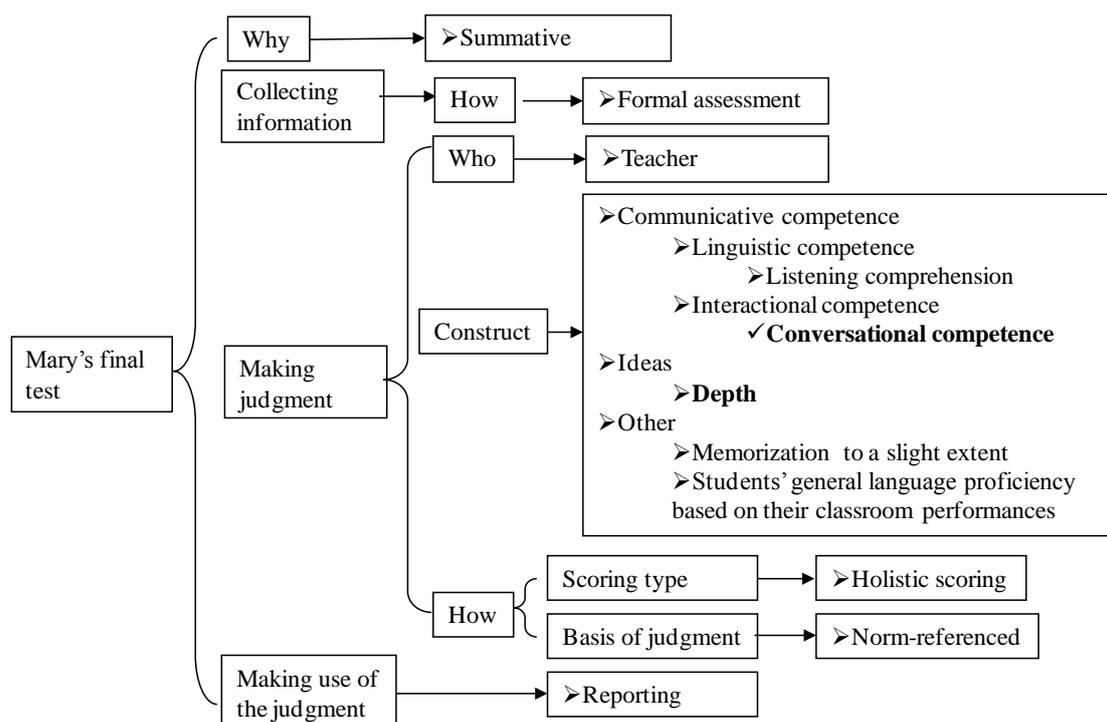
Figure 5.3 Profile of Mary's Final Test

5.2.3.6 Students' perceptions

The students' end-of-semester interviews showed that while they all prepared for the final test, they were in general not that worried about the test, either before or during the test.

To prepare for the final test, Lucy watched the film again and translated some sentences of the speech she selected into Chinese; Jody worked out an outline about what she would say on each of the two topics and rehearsed by herself the night before the final test; Daniel spent one hour and a half discussing the two topics with his

partner two days before the final test; Vivian had been considering the two assigned topics whenever she had time during the week before the test.

At the time of the test, the students were a little nervous the moment before they entered the room for the test or at the beginning of the test, but once they got started, they were not that nervous.

5.2.4 Unrecognized CA

5.2.4.1 Data

The profile of Mary's unrecognized CA practices was generated from the detailed analysis of the classroom recording transcripts of three observed lessons (the third observed lesson was devoted to the midterm test) and the relevant data from participants' interviews and student journals. Table 5.8 provides a brief description of the classroom activities during the three observed lessons.

Table 5.8 Mary's Classroom Activities

Duration in class	Classroom activity	Brief description
Week 7		
8:10-8:36	Teacher lecture	Mary gave a presentation on Halloween with the help of some power point slides.
8:37-8:50	Vocabulary activity*	Topic: Words about supernatural or unnatural creatures and people Format: Students working in groups of three for eight minutes before a teacher-led class discussion on the words they had found
8:51-9:06	Teacher sharing	Mary showed some pictures of herself and her family and friends, taken during previous Halloween holidays, and told students the stories behind those pictures.
9:07-9:08	Teacher instruction on the homework	Mary assigned the homework: Prepare to tell a scary story.
9:09-9:19	Teacher sharing	Mary read aloud a poem *The Raven* by Edgar Allen Poe to the whole class.
9:20-9:25 Break		
9:26-9:55	Class discussion*	Topic: What do you think of your own family members? Source: Assignment from the previous lesson Format: Mary provided a list of words relating to the discussion topic for two minutes. Then students talked in pairs for 12 minutes. Then Mary conducted a class discussion for 15 minutes allowing nine students to

		share their ideas.
Week 8		
8:00-8:20	Warming-up chatting	Four students talked about what they did during the past weekend and Mary provided online feedback for each student's talk.
8:21-9:55	Story telling*	Topic: Tell a scary story Source: Assignment from the previous lesson Format: Student worked in pairs first for nine minutes and then Mary asked 12 students to tell their scary stories in turn and she provided online feedback for each student. There was a 12-minute break in between.
Week 15		
8:00-8:08	Warming-up chatting	Mary had a chat with some students about the weather and about getting a cold.
8:09-8:37	Watching movie	Mary played the second half of the cartoon movie *Christmas Carol*, which they did not watch during the previous lesson.
8:38-9:14	Class discussion*	Topic: Message of the story *Christmas Carol* Source: Assignment from the previous week Format: First it was a teacher-led group retelling of the first half of the movie because some students were absent for the previous lesson. Then it was a teacher-led class discussion on the message of the story.
9:15-9:25 Break		
9:26-9:37	Teacher instruction about the final test	Mary informed students about the time, place, format, and content of the final test and then answered questions from students about the final test.
9:38-9:58	Teacher lecture	Mary gave a presentation on Christmas with the help of some power point slides.

It can be seen from the above table that Mary conducted a variety of classroom activities in her class. Analysis showed that four of Mary's classroom activities (marked with a * in Table 5.8) involved assessment: the story-telling activity, the two discussion activities, and the vocabulary activity. The first three activities were all assignments from their respective previous weeks and all involved students' performances and Mary's immediate feedback based on the students' performances. As to the vocabulary activity, though no evidence was found that this was also an assignment from the previous week, Mary did elicit students' performances by assigning them a vocabulary task and then provided feedback based on the words students had found, which showed the three steps of CA. In addition, analysis also found many *incidental assessment episodes* occurring during the feedback stage of the above-mentioned four activities, as well as during Mary's lectures and the warming-up chatting period of the week 8 lesson. The following table shows the percentage of such assessment-involving activities and episodes in the total observed class hours.

Table 5.9 Overview of Mary's Classroom Activities

Classroom activity	Frequency of occurrence	Total amount of class time (minutes)	Percentage in the total observed class time
Teacher lecturing	2	48	16%
*Class discussion	2	67	21%
Warming-up chatting	1	20	7%
*Story telling	1	83	27%
*Vocabulary activity	1	14	5%

It can be seen that although the total number of Mary's assessment-involving activities was not as large as that of Andrew's (cf. Table 4.10), over half of her class time was also engaged in assessment-involving activities, though she did not think she was assessing her students (Mary_EoSI). In what follows, these identified CA practices will be described in detail.

5.2.4.2 Unrecognized CA practices

It should be pointed out that this group of activities showed the following common features.

First, analysis showed that her purpose of conducting such assessment-involving classroom activities was primarily *formative*. For one thing, Mary intended her classroom activities to be learning or practicing opportunities for students, because she thought her primary job at class was "to get my students to talk" (Mary_BI, Mary_EoSI). For another, by engaging students in extended conversations during her classroom teaching, she was able to find out her students' strengths and weaknesses, which helped her adjust her teaching. "If you notice what people need, then you try to design some kind of activity that's gonna help them, meet that need" (Mary_EoSI). Moreover, she often provided online help as students talked because she want to promote learning.

> I noticed that if I put them in groups, the conversation is often rather dry. It can be, sometimes, depends on the students, but if I get into the middle of the conversation, it tends to be a little more lively, or maybe I challenge them to think other thoughts than they are used to. (Mary_EoSI)

While her purpose was primarily formative, she did form general impressions about her students' general language proficiency, which in turn influenced her final grading (cf. Section 5.2.1). This reflected the fact that to some extent Mary's unrecognized CA practices also fulfilled a *summative* purpose.

Besides, regarding the *who* component of the *making judgment* step, classroom observation revealed that Mary had always been the major agent of judgment and only occasionally some students would help a classmate with some correct expressions or pronunciations (e.g., turns 9 & 13 in Excerpt 5.2).

Moreover, the *how* component of the *making judgment* step in the descriptive framework (Figure 3.3) was found not quite relevant for this group of instruction-embedded CA practices.

Finally, regarding the *making use of the judgment* step, analysis revealed that Mary often provided detailed feedback, most of which are task-referenced. However, occasionally, there was also person-referenced feedback, such as "Good", "That's great!" In addition, Mary's judgment was also constantly used to inform her *teaching*, as can be seen from the following quote. When being asked why she spent so much time doing the story-telling task in the week 8 lesson, Mary explained that

> You know I kind of have to go by what's going on. You know you never quite know how it's gonna go, you know whether people have a lot to say or a little to say, yeah, so I always have something else to share, to talk, to go on to. You know sometimes you have classes very very quiet and don't have much to say, and you have to have a lot of other things to back it up with. (Mary_SRI)

Clearly, her classroom teaching evolved from her constant assessment of her students' performances in class. In addition, as mentioned above, the fact that the general impression she got about her students during such CA-involving activities influenced her final grading indicated that her assessment was also used for *reporting* purposes.

Considering such commonalities and to avoid repetition, the description of this group of unrecognized CA practices will focus on the *assessment method* component and the *construct* component of the *making judgment* step. This part will first describe the four planned assessment-involving instructional activities together with the incidental

assessment episodes contained within each, and then describe the incidental assessment episodes identified in teacher lectures and one warming-up chatting period. Students' perceptions of this group of practices will be presented after the descriptions of all the unrecognized CA practices.

CA practices within the vocabulary activity

Mary conducted one vocabulary activity in her week 7 lesson. She asked students to come up with words relating to supernatural things and classify them into groups. After students worked in groups of three for eight minutes, she first asked the total number of words each group had got, and then asked one student from each group to share with the class about the words they had found and gave her confirmative and/or explanatory feedback. Excerpt 5.1 is taken from the feedback stage when Mary asked one student to share with the class about what words her group had found.

Excerpt 5.1

Turn	Speaker	Transcript	Feedback type
1	T	Well, let's start from the smallest group first. So seven, you have seven?	
2	SS	Yeah.	
3	T	Ok. Let's hear which ones you have. ((Mary indicated one student to talk.))	
4	S1	Zombie.	
5	T	Oh, zombie, yeah, yeah, yeah, ok. Got it, got it. What else?	Confirmative
6	S1	Superman.	
7	T	Superman, ok. Superman is supernatural, yeah, power, ok. What's next?	Confirmative Explanatory
8	S1	Spiderman.	
9	T	Spiderman, oh, every good boy's hero, yeah? Spiderman ok.	Confirmative Explanatory
[…]	[…]		
18	S1	Shrek.	
19	T	What's Shrek called? What kind of monster is he?	
20	S1	Monster.	
21	S2	A green monster?	
22	S3	No no no, it's ... or something.	
23	S2	Oh, ogre	
24	T	Yeah, ogre	Confirmative
25	SS	Ogre, ogre, ogre ((Some students were looking up the word in their e-dictionaries.))	
26	T	Yeah, it's an ogre. Good. ok, what else?	Confirmative

Analysis showed that the whole activity was a *planned assessment-involving instructional activity* because Mary had planned this activity in advance since she had already put the task in her power point slide. However, within this activity, there were many *incidental assessment episodes* centring on the words / expressions students provided (e.g., turns 3-5, 5-7, 7-9), and these episodes were in the form of IRF sequence (Sinclair & Coulthard, 1975). Because these words were not planned beforehand and Mary's feedback was contingent upon students' actual performances, these episodes were incidental in nature. Besides, Mary mainly provided confirmative feedback (21 instances) and explanatory feedback (seven instances), and there was only one instance of corrective feedback in this activity.

Obviously, the focus of this activity was students' vocabulary knowledge, a kind of students' *linguistic competence*.

CA practices within the story-telling activity

In her week 8 lesson, Mary spent almost three quarters of the class time doing the story-telling activity. She asked students to practice in groups first and then asked individual students to share their stories with the whole class. When one student was talking, Mary often provided online comments and some corrections and suggestions. After the student finished his/her story, Mary just gave a brief comment and then asked another student to tell his/her story. This was an assignment-checking activity and Mary provided feedback based on students' performances. Therefore, the whole activity was a planned assessment-involving instructional activity.

Further analysis revealed that within this activity there were 12 student-level planned assessment-involving instructional episodes, because 12 students were asked to share their stories with the whole class. Each student-level CA practice usually started with Mary's indicating some student to tell his/her stories and ended with a general comment and an invitation to another student. Excerpt 5.2 is about one student's story-telling. However, within a student-level CA episode, there were also 51 incidental assessment episodes each centring around one specific language point (e.g., turns 6-7, 8-10, 10-12, 12-13, 14-15, 16-18, in Excerpt 5.2). They were incidental because Mary's feedback was contingent upon students' actual performances.

164

Therefore, an embedded structure was revealed with incidental assessment episodes embedded within planned student-level assessment-involving instructional episodes, which were further embedded within this planned assessment-involving activity.

Excerpt 5.2

Turn	Speaker	Transcript	Feedback type
1	T	She appeared as his girlfriend, oh wow! ((This is Mary's general comment on the previous student's story.)) Ok, somebody else? Ok, Mark.	
2	S1	I remember a story, a Chinese story just now. It's scary, so I want to turn off the lights.	
3	T	Really? I don't know whether we are ready for that. ((Mary and the other students laugh. One student turns off the lights.))	
4	S1	This is a Chinese scary story, so I wish I wish you can help me to translate it. ((He is talking to his classmates.))	
5	T/SS	Ok.	
6	S1	Er, one day, … a driver was driving was driving from countryside to town. It's a long road. Er ... so er... the the there is no one on on on the roadside. There is just just a sun in the sky so it's scaring, but but but the driver don't afraid of of it.	
7	T	<u>is</u> not afraid of it	Corrective
8	S1	he is listen to music rock music and yeah he is listen to rock music but suddenly suddenly er … there is an old woman in front in front of the, … on the roadside on the roadside, and and she 招手 ((Zhao Shou, waved))	
9	S2	Waved	Corrective
10	S1	And she waved. So the driver er... slow down and pull out	
11	T	Pulled over	Corrective
12	S1	Pulled over and the old the old woman said would you mind me, would you mind taking me to the torn?	
13	SS	Town, town	Corrective
14	S1	The driver was helpful so he don't afraid of her	
15	T	He <u>wasn't, wasn't</u> afraid of her.	Corrective
16	S1	And and he and he let him er... get off the car and and they and er when the driver er...get to the town er he he he turned around to see the old woman, but he don't see.	
17	T	didn't, didn't see her	Corrective
18	S1	He didn't see him see her, so he is puzzled and er […]	
		[…] ((S1 finished his story. There were four other instances when Mary and some students helped him by providing three correct expressions and one correct pronunciation. There were still many other grammatical and pronunciation errors that were left uncorrected.))	
36	S1	So if you walk around the roadside at night, you must be careful. Maybe something is following you.	
37	T	Hahaha ... ok. Thank you for your advice. Hahaha...oh good! oh! All right, Annie, I think we have time for one more.	

Concerning the assessment constructs, analysis revealed that Mary was primarily concerned with students' linguistic competence, as can be seen from Table 5.10, and her feedback was primarily corrective (49 instances).

Table 5.10 Aspects Mentioned during Mary's Story-Telling Activity

Aspect	Frequency of occurrence	Example
Linguistic competence (51)		
Providing a needed word or expression	19	Bunk beds, mannequin, vampire, hypnosis, put forward, etc.
Correcting an inappropriate word or expression	14	Took down her clothes and shoes → took off Fearly → fearful Before → ahead of time Make them all dead → kill them all Sympathy → sympathetic Etc.
Grammar	14	His → her Wear → was wearing To be die → to die Stands → standing Don't → didn't Etc.
Pronunciation	4	Hole (hall), monster (master), slipped (slept), bomb (boom)
Sociocultural competence	1	M: Well, better if you try to tell the story, not to read it. …Do you know <u>the difference between written and spoken</u>? Yeah, written story has a lot of descriptions that you know and the story normal is very concise, yeah, so it makes a meaning, yeah, makes a meaning. So when you tell a story you have to change it somewhat, yeah, it's a little different. I tell to the third person. You know third person? Yeah, the story was written for the first person, but you are gonna tell the story what happened by probably changing it to the third person. It was a story about a guy. He was in a he was in school and it was a library. So that's you know if you want to tell it orally, then you make that change, right?
Student attitude	1	M: Ok Steve, why don't you start? So can you tell us your story? S1: Sorry, I have I have I have I have no story. M: You didn't bring your story? You don't know any scary stories? S1: Er...actually I have scary story but I can't retell it in English so M: Oh, you were supposed to practice it this week so you could tell it in English. That was the assignment, yeah. You should have practiced so you can bring your story and retell it. Next time, next time, ok.

It should be pointed out that Mary's online feedback on students' linguistic competence was actually a form of scaffolding, which, in the field of second language learning, is described as the language an interlocutor uses to support the communicative success of another speaker including the provision of missing vocabulary or the expansion of the speaker's incomplete sentence (Lightbown & Spada, 2006, p. 204). That Mary's feedback was primarily to correct linguistic forms revealed that Mary's scaffolding strategy was mainly "modelling potential answers" (Booth, 2012, pp. 19-21). Since FA and scaffolding are essentially the same thing in light of sociocultural learning theory and Vygotsky's zone of proximal development (Shepard, 2005), Mary was obviously engaged in FA while conducting this activity.

CA practices within the discussion activities

The dataset for the present study contained two discussion activities, both of which were assignment-checking activities. Therefore, both were planned assessment-involving instructional activities. Analysis showed that both activities contained student-level assessment episodes during which Mary interacted with individual students. Within such student-level assessment episodes, there were many incidental assessment episodes where Mary mainly assessed the specific ideas students had expressed and pushed students to talk more and talk more deeply by asking specific contextualized questions or making contextualized comments, a strategy similar to Dawn's "contextualizing assessment content for students" (Booth, 2012, pp. 16-17). Such features can be seen from the following two excerpts. Those turns that contain such scaffolding questions and/or comments are marked with a * after the numbers of those turns. In the excerpts in this thesis, if a turn number is marked with a *, it means that turn contains scaffolding questions and/or comments.

Excerpt 5.3 is taken from the week 7 lesson in which Mary asked students to share their opinions on their family members, especially their parents and grandparents. In this excerpt, Mary was helping one student to express her opinions. Excerpt 5.4 is taken from the week 15 lesson and the discussion topic was what was the message of the film: *The Christmas Carol*.

Excerpt 5.3

Turn	Speaker	Transcript	Feedback type
1	T	Yeah, so, what do you think? Were your parents and grandparents overindulgent?	
2	S1	Maybe my grandparents, both my mother's mother and my pa my path's mother	
3	T	Father, father's mother	Corrective
4	S1	father's mother	
5	T	Yeah.	
6	S1	When when I came to er because they didn't live with us, so when I come when I go to their homes, they just eh...拉到我，把她拉到我这边怎么说？((La Dao Wo, Ba Ta La Dao Wo Zhe Bian Zen Me Shuo? Pull me, how to say "pull me to her side" in English?))	
7	S2	Pull you.	
8	S1	Oh, pull, they will pull me to their side and ask me "what do you want? Do you need more clothes? Or do you need some"?	
9*	T	Really? So they just wanna buy things for you?	
10	S1	Yeah, and every time I came to go to their house they will give me some money in the <xxx>, but I don't really so …	
11*	T	You didn't like that?	
12	S1	No, I like that. But er...because my parents, my parents are near, so I tries to be more mature.	
13*	T	Mature. Did you say you didn't want to, or they didn't mean to buy things for you?	
14	S1	Actually I very want.	
15*	T	You really wanted it, you really wanted it, yeah? Maybe you should not, you shouldn't accept it, you should refuse it yeah?	Corrective
16	S1	Yeah, I should say that "No, grandma, I I already have a lot, you don't have to".	
17	T	Yeah.	
18	S1	So when I said, I fell very sorry.	
19*	T	Very sorry. Did they not give it to you then?	
20	S1	Maybe sometimes my parents will go out, and will, they will give me.	
21	T	Hahaha... they gave it to you anyway, oh, that's interesting. Ok, anybody else?	

Excerpt 5.4

Turn	Speaker	Transcript	Feedback type
1	T	Yeah, that's right. What time is it? Ok, somebody else, Jim, I haven't heard from you.	
2	S1	Me? ((Jim had been looking down reading something and looked a little startled when hearing his name called.))	
3	T	Yeah. ((The class laugh.)) Are you reading Dickens there I hope? I hope you are reading Charles Dickens.	

4	S1	Uh, from this movie I can know that the childhood is very important to one's, childhood is very important to someone. It can, what he experience when he was young can have a great effect to to his, when he grew old and when he backs and his behaviors, his attitudes towards life. So this this character in this movie is influenced by his childhood experience, adapted experience, someone left her and abandoned her and	
5	T	Him , yeah.	Corrective
6	S1	No one spent Christmas with him. So uh I love this story that uh if we are, uh, we should uh treat a child with our honest heart and do not leave a bad impression and don't, so if there are, if it really happens in our in our daily life, in our real life, so one can't one can't became turn over turn into uh someone searching for gold and trying to make money, can turn him into a good mind. The man, just one night, so that happened.	
7*	T	Um, ok, well, I might disagree with you, but anyway, it would take a miracle to change somebody. Yeah, but, er but do you think that's the message of the story though? That if you have a bad childhood, then you're going to turn out bad?	Explanatory
8	S1	In other respect.	
9	T	Sorry?	
10	S1	In other respect.	
11*	T	In some respect, well, it does, obviously does show that, but you know, because today, in today's world, I know especially in America, a lot of people blame their present situation on their past, you know, like well, my mother had treated me differently. If my dad had paid attention to me, if I hadn't grown up in this circumstance, but is that the message here? What is the real message here though?	Corrective Explanatory
...	...	[...] ((Mary looked around and a girl gave an answer. She talked about the message of the whole story rather than Scrooge's childhood. Mary continued to invite other students to express their ideas. Two other students attempted an answer but what they said to some extent diverted from the original question. After acknowledging what each had said, Mary gave the following feedback.))	
18	T	Yeah, so anybody else? Comment? But I'm going to go back to this idea, though, his childhood, because I think you know when I see him, there are stages in his life to make choices, right? You know somebody in here talked about Nick. Did somebody do his speech? Nick the guy with no arms and no legs? In one of the classes, someone brought in the speech by him and showed a video. Anyway do you know what I am talking about? He was here in Beijing about a couple of weeks ago? Anyway, born with no arms and no legs, and if you look at his life, I mean he's obviously made the best of his circumstances. So we all have excuses or we make choices. So we can't, basically I think what the message here is that we can't use excuses for our behavior. Do you know what I mean? We all have choices to make. We can choose to love, to hate, to be greedy, or to give, you know, to let our past determine our future or to make choices where we change our future. You know what I mean? So I think that was a big message in the film.	Explanatory

It can be seen that Mary only occasionally corrected student's language errors (e.g., turn 3 in Excerpt 5.3, turn 5 in Excerpt 5.4). In contrast, the majority of her online feedback was in the form of questions, pushing students to clarify their ideas (turns 11, 13, 19 in Excerpt 5.3, turns 7 & 11 in Excerpt 5.4) or offering an expression that expressed the student's idea in a clearer and more precise way (turns 9 & 15 in Excerpt 5.3, turns 7 & 11 in Excerpt 5.4). Her way of conducting such discussion activities revealed the features of DA (dynamic assessment), especially interactionist DA (Poehner, 2008).

Table 5.11 presents the frequency counts of different types of feedback focus. It can be seen that linguistic competence was not a major focus during such discussion activities; instead, the clarity and depth of students' ideas received primary focus.

Table 5.11 Aspects Mentioned during Mary's Discussion Activities

Aspect	Frequency	Example
Linguistic competence (8)		
Grammar	4	Feel heart warm → feel quite warm in your heart Youngest children → youngest child Him → her Very want → really wanted
Providing a needed word	3	Advice, sibling, greedy
Pronunciation	1	Turn 3 in Excerpt 5.3
Ideas (50)		
Clarity	38	S1: My mother is supportive and patient, and I think because of that I'm a little independent. … I I I call her every day. T: You are dependent you mean? Not independent?
Depth	11	S1: He was very lively and he loves dancing very much. T: That's right ((laugh)). So he loved dancing. Ok. So what do you think was the lesson in there? I mean why did they show us that? Why did Dickens show us that scene?
Logic	1	Turn 7 in Excerpt 5.4

Incidental CA episodes within teacher lectures

In Mary's class, she sometimes would give a presentation on a topic with the help of some power point slides (Mary_COF26102010, Mary_COF28122010). While her lectures were clearly not CA practices, sometimes during her lectures she would pause

and ask the students a question about a difficult word or some cultural knowledge to check if they were with her, as can be seen from excerpt 5.5. These episodes generally took the typical IRF sequence (Sinclair & Coulthard, 1975) and the explicitness and comprehensiveness of her feedback varied according to students' responses. Therefore, such IRF sequences were regarded as assessment-involving episodes. Since such IRF sequences occurred as her lectures proceeded, they were incidental in nature.

Excerpt 5.5 is taken from Mary's presentation on Halloween in the week 7 lesson. Those parts in bold indicate Mary's eliciting steps, marking the beginning of an IRF sequence.

Excerpt 5.5

Turn	Speaker	Transcript	Feedback type
1	T	[…] So what do you know about Holloween?	
2	SS	-Pumpkins -Children	
3	T	Pumpkin, that's right.	Confirmative
4	SS	-Children -Trick or treat	
5	T	Sorry? trick or treat? Yeah, why do they say trick or treat?	
6	S1	Er... if you don't give me sweets, I will play tricks on you.	
7	T	That's right, yeah, trick, I mean that's true. Most kids don't play tricks, but some do, especially when they get older, actually little kids don't usually play tricks, but teenages do. They like this idea, you know, they get together and do some naughty things on Halloween. And **what else do you know?**	Confirmative
[…]	[…]		
24	T	... but today it's not a Christian holiday, it's a secular holiday. **Do you know secular? What does the word secular mean?**	
25	SS	((No response, in silence))	
26	T	Secular means not religious, ok, so you have religious, you have secular. So secular will be separated from anything religious, right? //Yeah, so, some Christians and pagans, **you know what pagan is? Who's a pagan?** You can maybe guess from: Christians and pagans.	Explanatory
27	S1	Someone who believes in God?	
28	T	No, someone who doesn't actually. A pagan is someone who is not a Christian, ok. So Christians are other people than pagans, right. So, but in the past, I think even today, pagan has the idea of being er very er… you know non-ethical person, you know. A pagan is someone who parties and drinks and carries on, does all kinds of crazy bad things, right? Yeah, but anyway, anyway, some	Corrective & Explanatory

		Christians and pagans like non-believers. A pagan would be a non-believer!	
[...]	[...]	[...]	
32	T	The Irish didn't have pumpkins, but they carve turnips. so turnips are big,... it's like potato type. You have turnips here, you know, big, er…, funny shape thing. They would carve those which probably is not easy to carve as a pumpkin, cos pumpkins are hollow. **Do you know hollow?**	
33	SS	Empty.	
34	T	Yeah, hollow is empty inside, so they were easy to cut. Anyway, so we often do that at Halloween, yeah. Here is a painting from the Irish celebration. [...]	Confirmative

It can be seen that there are five complete IRF sequences in this excerpt. For the first two, Mary checked students' background knowledge through questioning (turns 1 & 5) and then confirmed their correct answers (turns 3 & 7). For the other three, Mary checked if students knew the meaning of some difficult words through questioning (turns 24, 26, 32) and then provided feedback based on students' responses. For the word "hollow", when several students could provide a correct synonym of the word (turn 33), Mary simply confirmed their answer and carried on with her lecture. However, for the word "secular," no students attempted an explanation, and for the word "pagan", one student provided a wrong answer. Recognizing students' difficulties, Mary not only provided the correct meanings but also added an explanation. The variations relating to her feedback reflected that Mary adjusted her teaching based on her evaluation of students' relevant knowledge.

Altogether 16 instances of such kind of IRF episodes were identified from Mary's two lectures, one on Halloween in the week 7 lesson and one on Christmas in the week 15 lesson. While most of such episodes were about difficult vocabulary and cultural background knowledge, there was also one instance when Mary asked about students' past learning experiences and one about student's language use, as can be seen from the following table.

Table 5.12 Aspects mentioned during Mary's Lectures

Aspect	Frequency	Example
Linguistic competence (10)		
Vocabulary knowledge	9	Pagan, secular, harvest, haunting, abandon, cemetery, foretell, prophecy, hollow
Language use	1	S: [...] He is very clever so he invented it, to set signal to the army.

		T: Oh, yeah, that makes sense. You should say "to send signals to the army."
Cultural knowledge	6	What do you know about Halloween? Why do people say trick or treat? Why did people like to set off firecrackers in the past? Do you know Frankenstein? Do you know Mary, Virgin Mary? Do you know the two big sections of the Bible?
Student past learning experiences	1	How many of you have never heard of Halloween?

Incidental CA episodes within warming-up chatting

Mary usually chatted with her students at the beginning of a lesson to "warm students up" (Mary_SRI). Analysis revealed that during the chatting period of her week 8 lesson, there were many incidental assessment episodes. By raising scaffolding questions, making scaffolding comments, and providing corrective feedback, Mary helped four students talk about how they spent their previous weekend. Similar to how she conducted the story-telling activity and the discussion activities, Mary chatted with one student until that student finished his/her story before she moved on to talk with another student. Although Mary made no evaluative comments until one student finished his/her story by saying "very great, cool" or "that's fine, that's great experience", her online questions and comments reflected the fact that she assessed the comprehensibility of what the student said as the students talked about their experiences (see Excerpt 5.6). Those turns that contain scaffolding questions or comments are marked with a star after the numbers of those turns.

Excerpt 5.6 shows how Mary provided online feedback, including providing a needed or a better expression, correcting an error, asking scaffolding questions or making scaffolding comments, to help the student finish his story. Such online feedback showed that Mary was constantly assessing the idea and language of what the student said and was engaged in FA practices.

Excerpt 5.6

Turn	Speaker	Transcript	Feedback type
		((After helping one student finish her story about her weekend, Mary looked around to see if someone else had something interesting to tell.))	
1	S1	I had the more the worst weekend uh over the over the two	

		years of my life.	
2	T	Are you serious? ((Laugh)) / Oh, no, the worst weekend ever, uh, we will say the weekend from hell, yeah.	Corrective
3	S1	On Friday night, I went to karaoke with my friends, and when I left I forgot my my full yeah, my full left in Karaoke, [so	
4*	T	[Forgot your?	
5	S2	Cell phone.	Corrective
6	T	Oh your phone! Oh my God!	
7	S1	So I went back. Then when I got got into my dormitory, I realized that, realizing, and I went back to to get my phone.	
8	T	Yeah?	
9	S1	And when I there is there are stone bricks. I I was riding a bike, and there were stones	
10*	T	On the road?	
11	S1	Narrow stones narrow gap so so I cross to the the gap and my bike and I all threw threw away.	
12	T	Really? / Went flying, we say went flying.	Corrective
13	S1	So my bike was broken.	
14	T	Oh.	
15	S1	And I got my I got my ((pointing to his elbow))	
16*	T	Elbow?	Corrective
17	S1	Elbow and my my ((pointing to his knees))	
18*	T	Knees?	Corrective
19	S1	Knees yeah.	
20*	T	Scraped, scraped?	Corrective
21	S1	and my body hurt.	
22	T	((class laughing)) We should laugh that he was all right, yeah.	
23	S1	So I have to mend my mend my bicycle that cost me 30 yuan.	
24	T	Oh.	
25	S1	Then on Sunday evening	
26	T	That was not the end, hahaha.	
27	S1	I went to a an argument held by law school	
28	T	Oh, a debate, debate.	Corrective
...	...	[...] ((The student told another unlucky experience.))	
39	T	That happens, so we call that the weekend from hell. You probably have heard that is a day from hell, so uh sometimes we say, this is an old-fashioned phrase, but people may say I got up on the wrong side of the bed today; everything is going wrong. Got up on the wrong side of the bed. Nothing it matters what (.) but you know we just say, then later a more common, or more current phrase is I'm having a bad hair day. Yeah. Canadians and Americans they made this up.((She wrote down this phrase on the blackboard.)) Of course, boys don't usually do that. It's a girl's thing, a bad hair day, or the weekend or the day from hell. ((She wrote down this phrase on the blackboard.)) What was the other one I said? Oh, got up on the wrong side of the bed. ((She wrote down this phrase on the blackboard.)) Sometimes bad things can happen, uh. Nothing goes right. ((Then she told the unlucky experiences of one of her friends.))	Explanatory

The following table presents the frequency of Mary's assessment focuses during this warming-up chatting period. It can be seen that Mary focused on students' ideas without overlooking their language quality. However, her feedback was mainly about correcting students' language errors.

Table 5.13 Aspects Mentioned during Mary's Warming-up Chatting

Focus	Frequency of occurrence	Example
Linguistic competence (9)		
Providing a needed / better word / expression	6	Mid-term election, follow, elbow, knee, scraped, the weekend from hell
Correcting an expression	2	Argument → debate Youth League → Communist Youth League
Pronunciation	1	Turns 3-6 in Excerpt 5.7
Clarity of ideas (25)		
(Through scaffolding questions)	18	S1: […] he made a distinguished difference between China and Western countries. T: Oh, how did he do that? What did he say?
(Through scaffolding comments)	7	S1: […] but the Western people they had their own beliefs, so they don't take politics as a kind of faith, just something they do as a career. T: Yeah, or as your duty as a citizen.

5.2.4.3 Profile of Mary's unrecognized CA practices

The following figure summarizes the key features of Mary's unrecognized CA practices. The numbers after specific types of task-referenced feedback show the frequencies of respective types of feedback found in the data.

Why ──────────────────────────► ➢Formative

Collecting
information ──► How ──► ➢**Planned** assessment-involving instructional activity/episode
                        ➢Incidental assessment episode

                Who ──► ➢**Teacher**

Making
judgment

Mary's
unrecognized CA
practices

        Construct ──► ➢Communicative competence
                        ➢**Linguistic competence**
                            ✓**Vocabulary**
                            ✓**Grammar**
                            ✓**Pronunciation**
                        ➢Sociocultural competence
                      ➢**Idea**
                        ➢**Clarity**
                        ➢Depth
                        ➢Logic
                      ➢Other
                        ➢Cultural background knowledge
                        ➢Student attitude towards study
                        ➢Student past learning experience

Making use of
the judgment ──────► ➢Learning (feedback)
                        ✓Person-referenced
                        ✓Task-referenced
                            ●Corrective (75)
                            ●Confirmative (34)
                            ●Explanatory (22)
                      ➢Teaching
                      ➢Reporting

Figure 5.4 Profile of Mary's Unrecognized CA Practices

5.2.4.4 Students' perceptions

It can be seen from the above analysis that Mary's unrecognized CA practices were highly intertwined with her teaching. Although Mary did not think she was assessing her students during such practices, all the four volunteer students thought that they were being assessed during the above-mentioned classroom activities, especially during the moments when Mary was talking to individual students. In their end-of-semester interviews, when asked if Mary used some other ways to assess students besides the midterm test and the final test, they all mentioned students' in-class performances. For example,

W: How did your teacher work out the final score for each student?

J: I think one part comes from her general impression based on our in-class performance, one part comes from the midterm test, and one part comes from the final test. (Jody_EoSI)

When asked what in-class performance included, Jody said

For in-class performance, I think it's mainly a matter of attitude. So if you are active in class, probably that might affect in some way. But I don't think she particularly took notes about this. Oh, I remember for a couple of times she checked our attendance, because quite a few students didn't come on those days. (Jody_EoSI)

Daniel also said that except for the midterm test and the final test, " I think there is another part, a hidden part, that is our daily performance during those discussion periods. This must account for certain percentage" (Daniel_EoSI). Vivian also thought in-class performance would affect the final score.

W: In addition to the final test, did your teacher use some other ways to assess students?

V: Midterm test, and also the general impression based on your daily performance. For example, sometimes in class, she asked some questions, and some students would be very active and would try to grasp every chance to talk, but some student didn't want to talk. Then she would notice that and would ask him to talk and he couldn't say anything after a long time.

W: So do you think Mary was assessing students' such performances?

V: Yes, sure. (Vivian_EoSI)

In addition, an interesting phenomenon in this class was observed each time, that is, except for a couple of very quiet students, all the others would come to the classroom quite early and try to grab a seat near where Mary would sit. During the class discussion time, students were also very eager to take a chance to talk. However, they mainly talked to Mary. Sometimes, those who did not talk were not very attentive to other students' talk. For example, Vivian said that if a student's English was too poor, she would not listen to him/her talking (Vivian_SRI). Daniel said while a classmate was telling a very popular Chinese ghost story, he felt it quite boring and he began to play with his mobile phone instead (Daniel_SRI). Therefore, although Mary said she was not assessing her students during those class activities, students attached great importance to their own classroom performances, especially the moments when they talked to Mary in class.

Regarding students' attitudes towards those assessment-involving instructional activities, while Jody felt the discussion/sharing activities were especially beneficial

for her improvement, the other three students found such activities not that beneficial and attributed their improvement mainly to their after-class self-study.

Jody loved the describing-family-member activity a lot because she felt she was able to open her mouth and talk and she had tried her best to speak as fluently as she could. She also felt she had enlarged her vocabulary and practiced her thinking ability (Jody_J1). Of the week 15 lesson, Jody said she learned a lot from watching that movie and participating in the follow-up discussion on the message of the story.

> During the retelling period, I listened to other students' ideas. Sometimes Mary would correct some mistakes and I have learned some knowledge about the language. Later, when we were discussing the message of the film, Mary would tell us her opinions. Sometimes due to our English proficiency level, we couldn't express our ideas quite accurately, and then Mary would deepen our ideas, and I felt I have learned a lot from those moments. (Jody_SRI)

However, the other three students all felt this course was not substantial enough. In the end-of-semester interview, Daniel said after one semester, he could "speak more fluently, with more content, and I can usually find the words to express my ideas. Also now when I speak a Chinese word, I often want to say it in English again". However, he did not attribute his improvement to this oral English course because he thought "this oral English course just gave me a chance to talk to Mary" (Daniel_EoSI). He did not feel he had learned much from this course because "this course was basically a time for chatting. You don't have to prepare anything in advance. You don't have to prepare anything for it. We just come and sit here chatting" (Daniel_EoSI). Instead, he attributed his improvement to his after-class effort such as memorizing vocabulary, listening to BBC, and watching American TV series. He wished his department could find an English native speaker as a language partner for each English-major student in his department. When asked to what extent Mary's classroom activities such as group discussion/sharing periods helped him improve, he said "not much". He further explained that first "the time for each student to talk in class is rather limited", "since everyone should talk in class, you can't talk much", and second "when the topic was not that interesting, too superficial, too shallow" he did not have a desire to talk in class. He especially disliked the vocabulary activity when students were asked to think of words about supernatural things and felt it meaningless (Daniel_J1). However, he

enjoyed the moments when he was talking to Mary directly and Mary gave him individual feedback, but unfortunately, such chances of true communication were usually small.

Lucy's opinion was similar to Daniel's. While she enjoyed the class discussion periods, she did not have many chances to talk. She wished the class could be smaller. Since the middle of that semester, she had begun to practice her oral English with a foreign partner (Lucy_SRI). She also felt her real improvement came from her after-class effort. "Your chance to talk in class is very limited, and it's not that useful to me. The more important thing is how you use English after class and how you communicate with others" (Lucy_EoSI).

Vivian had very good oral English ability. She could talk quite fluently and she was very active in class. Her data showed she wished the course to be more demanding. During the semester, she joined a university club where she had a lot of opportunities to communicate with foreign students. She also had a chance to work as a volunteer student for 'Beijing Forum', during which many people from around the world came to discuss some environmental issues. Such experiences greatly enhanced her oral English ability, but also made her feel this course did not give her enough support or practice. She wrote in her journal,

> I've just come back from Beijing Forum. It was like a 4-day monster training camp. Through communication with foreigners, I feel we English-major students have many social responsibilities to shoulder. So for those social responsibilities, I think this oral English course is not effective at all. I think there is an urgent need for us to make more pre-class preparation, and in class we should discuss some serious issues and develop a kind of international vision. (Vivian_J2)

In the end-of-semester interview, Vivian said this course was too "relaxing", "no preview, no review, no textbook, … you just come and sit here chatting". She thought it should be more "intensive". When asked in what way the course could be more intensive, she said,

> For example, talk about more serious topics or some sensitive issues, and give each student more pressure. For example, after a student talks, he should make some improvement and next time he should do a little better. If every time he

stays at the same level, that's not ok. Teacher should make students feel nervous. (Vivian_EoSI)

On the whole, the students had ambivalent opinions towards the typical way Mary conducted her unrecognized CA practices, that is, engaging with individual students while at the same time providing online feedback. On the one hand, students enjoyed the authentic communication with Mary and her immediate feedback, but on the other hand, this practice was very time-consuming and consequently each student's practice time was then too short.

5.3 Students' general perceptions

The return rate for the end-of-semester questionnaire in Mary's class was not very good. Only 11 students returned their responses. Independent sample t-tests on part II of the beginning and the end-of-semester questionnaire data showed that none of the changes in the six aspects reached a statistically significant level, although students' effort and their performance approach goals dropped a lot over the semester (see Table 5.14). The fact that none of the six aspects reached significance was probably due to the small number of students who actually responded the questionnaire.

Table 5.14 Mary's Students' Self-reported Oral English Learning over the Semester

Aspect	At the beginning of the semester		At the end of the semester		Mean difference
	No. of students	Mean	No. of students	Mean	
Self-evaluation	18	19.56	11	21.00	+0.44
Mastery goals	18	13.67	11	13.27	-0.40
Performance approach goals	17	12.41	11	10.55	-1.86
Performance avoidance goals	18	4.83	11	5.82	+0.99
Anxiety	18	7.89	11	6.82	-0.93
Effort	17	14.24	11	12.27	-2.04

Regarding the third part of the end-of-the-semester questionnaire, the response rate was even lower, and inconsistency was prevalent between students' perceptions and the author's judgment based on the data obtained from the four observed lessons, as can be seen in Table 5.15.

Table 5.15 Students' Perceptions of Mary's CA Practices

Please circle the practices that your teacher has used in this oral English course.	Frequency					Information from classroom observation and interview
	Once the whole semester	Every month	Every unit	Every lesson	Every activity	
Explain the learning objectives of the whole course.	0	2	1	1	5	Seldom
Explain the learning objectives of a particular unit.	1	2	3	1	0	Seldom
Explain the learning objectives of a particular activity.	5	2	0	0	0	Seldom
Explain the connections between the learning objectives of the whole course, of a particular unit, and of a particular activity.	1	0	0	3	0	Seldom
Explain the skills to be grasped in this course.	3	3	2	0	1	Unclear
Organize activities for students to practice the skills to be grasped in this course.	4	5	0	0	0	Every lesson
Observe how well students have grasped the skills.	3	5	1	1	0	Every lesson
Ask questions to check how well students have learned.	4	3	0	1	0	Every lesson
Comment on students' performances after they finish a classroom activity.	5	2	1	1	1	Every lesson
Point out students' strengths and weaknesses after they finish a classroom activity.	5	1	1	2	1	Seldom
Tell students the correct answers after students finish a classroom activity.	2	3	0	0	0	Not relevant
Tell students how to do better next time after students finish a classroom activity.	4	3	1	0	0	Seldom
Praise those students who perform well in doing classroom activities.	1	2	1	0	0	Seldom
Criticize those students who do not try their best to do a classroom activity.	0	1	0	0	3	Once
Encourage students not to be afraid of making mistakes.	2	4	1	0	0	Seldom
Organize students to do self-assessment.	1	1	1	0	2	Seldom
Organize students to do peer-assessment.	0	1	2	2	1	Seldom
Assign after-class homework.	1	2	2	1	0	Every lesson
Check students' homework in class.	2	1	2	1	0	Every lesson
Explain to the students how they are assessed in this course.	0	0	0	4	6	Before the midterm and final tests but not in every lesson
Inform students about what they are expected to do at the course test(s).	0	0	1	2	3	Before the midterm and final tests but

						not in every lesson
Explain to students how they can get high scores in the course test(s).	0	0	0	2	4	Unclear

One possible reason for such discrepancy might be that the questionnaire was not appropriately designed or modified for this specific class, so that students might have found it difficult to make their choices. In light of this inappropriateness, no further analysis was conducted on the data concerning the usefulness of the selected practices. Instead, this questionnaire was further revised and then used in Linda's class.

5.4. Summary

In this oral English course, Mary conducted a variety of assessment practices throughout the semester. While she only regarded the midterm test and the final test as assessment, because she believed that an assessment should be a comparatively more formal task/activity during which the examiner could step aside and observe students' performances and grade accurately, she admitted that, by observing students' in-class performances, she formed impressionistic opinions about her students' general language proficiencies, which influenced her final grading. In other words, her FA and SA practices were not clearly separated from each other; instead, her FA practices contributed to her summative grading, though not to a great extent.

This implicit contribution of FA to SA was perceived by all the volunteer students. Although at the beginning of the semester Mary did not tell students how they were to be assessed in this course, all the volunteer students felt that she was assessing them during those classroom activities when they interacted with her. Classroom observation confirmed that most students cared a lot about the chances to talk to Mary in class and tried their best to grasp such chances but paid much less attention when other students were talking to Mary.

This connection between FA and SA was also supported by the fact that the impersonated-speech task and the final test both evolved naturally from her classroom teaching. The impersonated-speech task was a result after she noticed that pronunciation was a serious problem for many students in her class. After the midterm test, she provided extra materials for students to help them further improve their

pronunciation. She also used the final test as an opportunity for students to focus further on two issues they had not discussed sufficiently in class. Therefore, her SA and FA were closely knitted and fed into each other.

Teaching in a "low structure" context (Wette & Barkhuizen, 2009), Mary had a lot of freedom to decide what to assess and how to assess. Regarding assessment constructs, while Mary repeatedly mentioned that the primary goal for this oral English course was to enhance students' "fluency", or rather their confidence in speaking in English, her feedback throughout the semester was seldom about students' fluency. Instead, her feedback was mainly about students' language accuracy, especially pronunciation, vocabulary and grammar. Regarding her assessment methods, to be consistent with her own belief about oral English learning[9], her typical way of assessing students, especially during the CA-involving classroom activities, was through interacting with individual students on a topic each student has something to share and at the same time providing online scaffolding. Consequently, many incidental CA episodes were embedded in her classroom activities. This supports the finding that some teaching/learning activities are also assessment activities (Hill, 2012).

In all her CA practices, Mary had always been the major agent of judgment. She seldom invited her students to make comments after one student had shared his/her ideas with the class. In most cases, students were not involved in peer-assessment or self-assessment.

Regarding her scoring type in the formal assessment practices, Mary found analytical scoring troublesome and difficult to use, and found her holistic scoring consistent over time and therefore reliable in her opinion. Regarding her basis of judgment, being influenced by one university assessment policy where only 40% of the students in one class could get a score above 85, she mainly adopted NRM when grading her students.

---

[9] "I like that theory of language learning and language teaching, that is you use language to accomplish goals, and then you become more fluent, so I think that was my goal, to give them opportunities to keep using English in a real way. I think there is so much inside of them already stored up, you know, a lot of knowledge about English stored up, a lot of patterns, a lot of vocabulary, and it's just a matter of taking it out of the box and using it." (Mary_EoSI)

183

Regarding how she used her assessment results, except for the final test which Mary used to measure students' achievement only, all her other CA practices served dual purposes simultaneously. On the one hand, she used her assessment results formatively to diagnose students' language problems so as to modify her teaching to enhance student learning; on the other hand, she used her judgment summatively to determine students' general language proficiency, which played a role in determining a student's final composite score for this course.

The four volunteer students mentioned that how active they were in class was part of the course assessment and consequently they all tried to grasp every opportunity to talk in class. However, Mary's typical way of interacting with individual students over an extended period made students feel that they did not get enough opportunity to practice in class. Consequently, they felt the course was not substantial enough.

Chapter 6. Linda's CA Practices

This chapter describes the results of the data analysis for Linda's case. This chapter follows the same presentation structure as those for the previous two chapters. After the description of her teaching context, Linda's formal assessment structure was presented, followed by the profiles of both recognized and unrecognized CA practices. Students' general impressions of her CA practices as revealed from the students' questionnaire data are presented after all the profiles. This chapter ends with a summary of the typical features of her CA practices.

6.1 Linda's teaching context

6.1.1 Educational background and past teaching experiences

Linda, a Chinese female in her mid-forties, had a strong educational background in language teaching. She had a Master's degree in TESOL from an American university, and one of her courses was about language testing. At the time of the study, she was doing a PhD degree in Applied Linguistics, and her particular interest was second language acquisition. During her teaching, she had attended many conferences on foreign language teaching and learning as well as on-the-job training workshops provided by her department or some other educational organizations (Linda_BI).

Linda was very experienced in teaching oral English, especially to first-year English-major students at her department. She had over 20 years of teaching experience. She had taught many language skill courses such as reading, writing, speaking, listening, and pronunciation to both first-year and second-year students, but most of the time she was teaching oral English to first-year students.

Traditionally, the *Oral English* course was offered to freshmen and sophomores at her department. Though the courses had different teaching content and they became increasingly more challenging as students moved up, they were all called *Oral English*. The way to distinguish them was by specifying levels, such as *Year One Oral English*. Three years before the time of the present study, curriculum reform began at her department as a response to the changed expectations from the society as well as to the

changes in students' characteristics and needs. "The traditional conversational English was not challenging enough because students were only asked to do role plays and simple conversations" (Linda_BI). Therefore, a focus was specified for the *Oral English* course for each semester of the first two years, and *Public Speaking* became the focus of the *Oral English* course for Year One Semester Two.

Linda had been the coordinator for the *Oral English* course for many years before the reform. At the time of the study, she was the course leader for the Public Speaking course. There were eight parallel classes and five teachers, since some teachers taught two classes. As the coordinator, she was responsible for organizing regular meetings of all the teachers teaching this course to discuss the syllabus, ways to assess students, and ways to conduct each lesson.

6.1.2 Her university, her school, and her students

Linda was teaching at the School of English and International Studies (hereafter SEIS) of a key foreign language university in Beijing. She was teaching at the same university as Andrew, but at a different school. Her school (SEIS) was considered one of the best English-major programs in China. According to the information shown on the official website of this university, the undergraduate program of SEIS "strives to cultivate graduates with a high proficiency in English language skills, profound cultural literacy, well-balanced knowledge structure, outstanding abilities in learning, critical thinking, creating, cooperating, leadership, and a strong sense of social responsibility" (from the university website, accessed on Feb. 17, 2012).

To achieve such purposes, SEIS had set its curriculum in such a way that the first two years of students' undergraduate study were mainly devoted to comprehensive and systematic language skill training as well as to developing their interpersonal communication skills. The Public Speaking course, which Linda was teaching, was one of those compulsory language skill courses. When students reached the third year and the fourth year, their study focus was shifted to content courses such as English language and culture, social and cultural studies, international politics and economy, translation theories and practice, and international journalism and communication, although they still received systematic training on translation and interpretation.

Usually, for a language skill course, the class was small, with around 24 students in one class.

The present study was conducted in the *Public Speaking* course offered to first-year English-major students at SEIS. Linda was teaching this course to one of eight parallel classes. This class of 25 students, eight male and 17 female, were new to her. Ten students came from foreign-language high schools, where they had had systematic training in English, especially in oral communication. Three students had taken external standardized tests before the start of the present study: one student had taken TOEFL and got a score of 99 out of 120; one student had taken IELTS and got a score of 7.5; and a third student had passed College English Test Band 4 and Band 6. Seven students had experiences of travelling abroad, all for a period of less than a month.

Questionnaire data revealed that at the beginning of the semester, students were confident about their English abilities but they were not satisfied with their present oral English proficiency (Table 6.1). Similar to Andrew's and Mary's students, Linda's students also had high mastery goals and performance approach goals, and low performance avoidance goals (Table 6.2).

Table 6.1 Linda's Students' Self-Evaluation of their Oral English Ability at the Beginning of the Semester

Questionnaire item	Not true at all	Not really true	Partly true	Mostly true	Absolutely true	Mean
1. I am very confident when I speak English.	4 (16%)	1 (4%)	9 (36%)	9 (36%)	2 (8%)	3.16
2. I worry that native English speakers will find my oral English strange.	2 (8%)	10 (40%)	7 (28%)	4 (16%)	2 (8%)	2.76
3. My oral English is very fluent.	3 (12%)	7 (28%)	9 (36%)	6 (24%)	0	2.72
4. My poor oral English always makes me feel inferior to those who can speak fluent oral English.	6 (24%)	9 (36%)	3 (12%)	4 (16%)	3 (12%)	2.56
5. When I speak in English, I can always find the words to express my ideas.	1 (4%)	5 (20%)	9 (36%)	9 (36%)	1 (4%)	3.16
6. When I speak in English, I can say exactly what I want to say.	3 (12%)	5 (20%)	9 (36%)	7 (28%)	0	2.83
7. I am satisfied with my present oral English proficiency.	5 (20%)	9 (36%)	11 (44%)	0	0	2.24

Table 6.2 Linda's Students' Goal Orientations at the Beginning of the Semester

Questionnaire item	Not true at all	Not really true	Partly true	Mostly true	Absolutely true	Mean
Mastery goals						
8. It's very important for me to completely grasp the skills taught in this course.	0	0	0	5 (20%)	20 (80%)	4.8
12. I am willing to spend a lot of time practicing to improve my oral English.	0	1 (4%)	1 (4%)	9 (36%)	14 (56%)	4.44
15. I believe that if I work hard, my oral English will improve.	0	0	0	13 (52%)	12 (48%)	4.48
Performance approach goals						
9. It is very important for me to be the top student in this oral English class.	0	2 (8%)	6 (24%)	14 (56%)	3 (12%)	3.72
11. I am eager to become a smart student in my teacher's eye in this course.	0	0	3 (12%)	13 (52%)	9 (36%)	4.24
14. I wish to get a high score in this course.	0	0	1 (4%)	16 (64%)	8 (32%)	4.28
Performance avoidance goals						
10. I will be satisfied so long as I can pass this course.	9 (36%)	9 (36%)	5 (20%)	2 (8%)	0	2
13. I worry that I may fail this course.	5 (20%)	12 (48%)	6 (24%)	0	2 (8%)	2.28
16. I feel that no matter how hard I try, my oral English will make little improvement.	12 (48%)	11 (44 %)	2 (8%)	0	0	1.6

Moreover, students' anxiety levels varied but not very high (Table 6.3), and most students were willing to put in much effort to improve their oral English (Table 6.4).

Table 6.3 Linda's Students' Anxiety Level at the Beginning of the Semester

Questionnaire item	Not true at all	Not really true	Partly true	Mostly true	Absolutely true	Mean
17. I get very nervous when a foreigner asks me something in English.	3 (12%)	1 (4%)	5 (20%)	10 (40%)	5 (20%)	3.54
18. I feel distressed about being unable to improve my oral English.	8 (32%)	10 (40%)	3 (12%)	3 (12%)	1 (4%)	2.16
19. I worry that other students will laugh at me when I speak English.	6 (24%)	10 (40%)	6 (24%)	1 (4%)	2 (8%)	2.32

Table 6.4 Linda's Students' Effort at the Beginning of the Semester

Questionnaire item	Not true at all	Not really true	Partly true	Mostly true	Absolutely true	Mean
20. I actively create opportunities to talk with others in English.	2 (8%)	4 (16%)	11 (44%)	7 (28%)	1 (4%)	3.04
21. I regularly enlarge my oral English vocabulary.	0	3 (12%)	6 (24%)	13 (52%)	3 (12%)	3.64
22. I try my best to improve the grammatical accuracy of my oral English.	0	0	4 (16%)	17 (68%)	4 (16%)	4
23. I grasp every opportunity to practice my oral English.	0	1 (4%)	8 (32%)	13 (52%)	3 (12%)	3.72

6.1.3 The oral English course

The oral English course for the first-year semester-two English-major students was called *Public Speaking*. For this 16-week course, the class met twice a week, one hour on Monday morning for the first lesson and two hours on Thursday morning for the second lesson. The course objective was "to cultivate the students' ability to speak effectively in public, with a clear sense of purpose, resourceful thinking, and confidence to express ideas" (Appendix 21). In the baseline interview, Linda also hoped that in addition to developing students' public speaking skills, this course "might also help improve their language quality", but she was not confident about that (Linda_BI).

In this course, two textbooks were used. The course was basically arranged around the public speaking skills explained in the book *The Art of Public Speaking* (Lucas, 2010), with each week covering one or two chapters of the book. However, not all the chapters were covered due to the limited time of the semester. Some chapters that were not quite relevant to students' needs at that time, such as "Analyzing Audience", were not included in the syllabus. The other textbook *Contemporary College English: Oral English* (2) (Yang, 2005) was used as "a resource book," as Linda explained that:

> It would be quite boring for me to explain all those skills all the time. So after I have explained some skills to them, I want them to practice using those skills. Then what should they talk about? If they could come up with some interesting and original ideas, that's great. But what if some students didn't know what to talk about? Then this book could just come in and serve as a kind of resource

book. Students could get some ideas and some language input from that book. (Linda_BI)

At the time of the study, it was Linda's third round to teach this course. As the course coordinator, she and her team members made changes to the course syllabus based on the teaching experience of the previous two rounds. They used to ask students to give five prepared speeches that would be marked, evenly spread out throughout the semester, and for the final test, they used to give students three topics to prepare and ask students to give a speech on one of the three topics at the test time. Then she and her colleagues found "the course was quite rushed", so they decided to "relax a little bit" this time (Linda_BI). As shown in their course description (Appendix 21), which was worked out by the teaching group together before the semester started, they decided to ask students to do three instead of five prepared speeches that would be marked during the latter half of the semester, and for the final test, they planned to have students give an impromptu speech instead of a prepared speech (Linda_BI). Such changes reflected the fact that Linda's teaching environment was towards the "low structure" end of the scale (Wette & Barkhuizen, 2009) because she and her colleagues had the freedom to negotiate the course syllabus rather than being given an externally specified syllabus.

This course was important for the students because it was a compulsory course and it would affect the comprehensive evaluation of a student at the end of the semester, which would determine if he or she could win a scholarship.

6.2 Linda's CA practices

6.2.1 Formal assessment structure

According to the course description (Appendix 21), student assessment in this course consisted of four parts: class participation (10%), making three prepared speeches (20% each), self-critique and reflection (10%), and the final exam (20%). Within class participation, attendance and interaction accounted for 5% and peer evaluation 5%.

Classroom observation and all the relevant interview and journal data showed that Linda followed this assessment plan. Specifically, she marked three prepared speeches

from each student during the semester, that is, an informative speech at weeks 6-7, and two persuasive speeches at weeks 9-10 and weeks 13-14 respectively. During the assessment of each prepared speech, she required students to do peer evaluation, both in marking each other's performances and in giving oral feedback. After each speech, students also had to write a self-critique or self-reflection journal. She also kept a record of student attendance. In week 16, students had their final test. All of these assessment practices conformed to the assessment plan.

Interview data showed that a student's final score for this course mainly came from their marks on the three prepared speeches and the final test. She used the other parts—class participation and self-critique—mainly for washback purposes rather than for assessment purposes, because they were designed to motivate positive learning attitudes and behaviours or to enhance student learning rather than assessing their learning, as can be seen from the following quotes.

Linda thought to assign 5% for class attendance was "to get across the message that it is a required course and class attendance is mandatory." Though this 5% also included "interaction", she did not really look at it, but putting it down in the assessment plan was to let the students know that

> This is a learning community, in which they need to communicate and learn from each other. They can't just bring their ears to the class. We want them to participate in class discussion and really listen and give feedback to their classmates. And we want them to interact with their peers not only with the class teacher. (Linda_EoSI)

For peer evaluation, Linda "didn't really evaluate the quality of their comments as long as they participated" (Linda_EoSI). Instead, she wanted to transmit to the students the following messages:

> One is to provide the speaker with immediate feedback from the listeners, and second is to consolidate their own understanding of public speaking skills and what is a good speech, and also to provide a chance for them to learn from each other. Probably it can also be related to fostering their critical thinking. (Linda_EoSI)

Besides, although she organized many classroom activities, during some of which she also encouraged her students to give peer feedback before she gave her own comments, students' performances during such peer evaluation periods were not assessed either.

Therefore, all her students could get the full 10% for class participation because 1) "no student was absent out of no reason", 2) they had all participated in the peer evaluation, and 3) none of the peer marks were "unreasonably high or unreasonably low" (Linda_EoSI).

For self-critique, "again I [Linda] didn't really give them a score for their critique journals," but "just want to give them a chance to reflect on their own performances. As long as they did that, it would be Ok" (Linda_EoSI). At the time of the end-of-semester interview with Linda, some students had not submitted their self-critique journals yet, so "today I [Linda] emailed them again saying that ok do you want this 10%. If you do, you'd better do" (Linda_EoSI). In other words, it was the completion of the self-critique journals rather than the quality of the critique that would ensure a full mark.

Figure 6.1 shows Linda's formal assessment structure. The dotted parts indicate the aspects that were assessed not based on quality but on completion. For these parts, almost every student got a full mark. They functioned mainly as a discipline tool and motivating tool. Therefore, students' final composite scores for this course were not totally about student achievement.
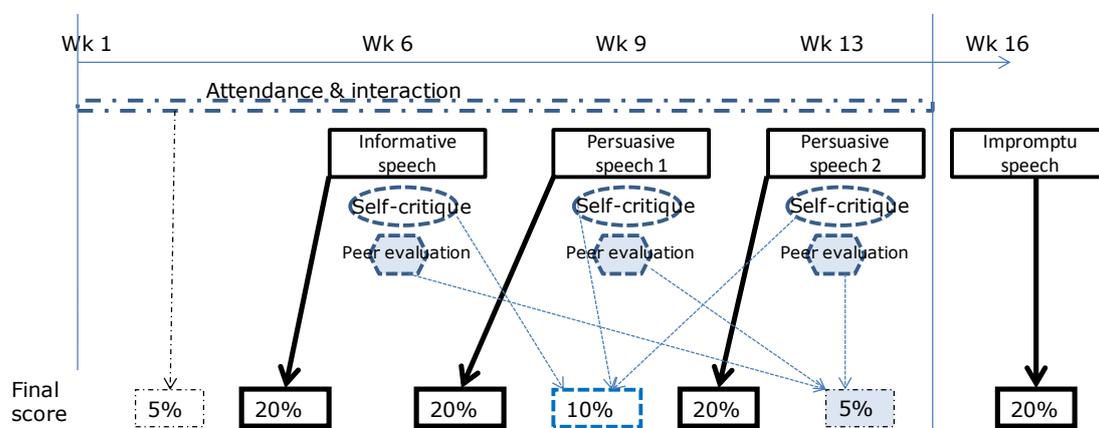
Figure 6.1 The Formal Assessment Structure in Linda's Public Speaking Course

Classroom observation revealed that some unrecognized CA practices occurred within some of Linda's classroom activities, though Linda did not regard them as assessment or use them to grade students. In what follows, the profile of the prepared-speech tasks and the final exam will be presented before the description of those unrecognized CA practices.

6.2.2 The prepared-speech assignments

Linda asked students to do three prepared speeches altogether. The first one was an informative speech and the other two were persuasive speeches. To generate a profile of this set of tasks, the author analyzed the relevant documents (Appendixes 21 to 24), the classroom recording transcripts and the observation field notes of the week 9 and week 13 lessons, as well as the relevant parts in students' journals, student interviews and the teacher interviews. Using the framework outlined in chapter 3 (Figure 3.3), these prepared-speech assignments are described from the following four aspects: assessment purpose, information-collecting method, judgment-making process, and making use of the judgment. Students' perceptions are described after the profile of this set of tasks.

6.2.2.1 Assessment purpose

Analysis showed that this set of tasks served both summative and formative purposes. They were for summative purpose out of the following three reasons. First, the three prepared-speech assignments accounted for 60% of the final grading. Second, the way they were arranged showed that they were designed for assessing student achievement. Analysis of the course schedule (Appendix 21) and the textbooks revealed that each prepared speech was arranged after students had learned all the necessary skills to make that kind of speech. Specifically, the first prepared speech, an informative speech, was arranged after students had learned how to select a topic (week 3), how to organize a speech (week 4), how to inform (week 5), and how to support one's ideas (weeks 3&5). The second prepared speech, a persuasive speech, was arranged two weeks later after students had learned how to persuade (week 7) and the various methods of persuasion (week 8). The third prepared speech, another persuasive speech, was arranged another two weeks later, during which Linda re-emphasized the delivery

skills (week 10) and how to outline a speech (week 11). Such an arrangement revealed that these prepared-speech tasks were designed to see how well students had learned the public speaking skills, and thus summative in nature. Third, the written instructions for the two persuasive speech assignments (Appendix 22) confirmed the summative purpose of this set of tasks. These two prepared speeches were opportunities for students to "apply the principles of speech organization, delivery, and persuasion ... covered in your readings and/or class lectures to date." Therefore, these assignments were specifically intended to assess what students had achieved in this course. This was also consistent with Linda's belief that it is "totally unfair" to assess such productive skills like speaking and writing only at the end of the semester, but such skills should be assessed systematically throughout a whole semester (Linda_EoSI).

Besides the summative purpose, the way the three assignments were arranged also showed the formative purposes behind their design. First, the students were informed of the three assignments at the beginning of the semester through Linda's introduction to the course (Linda_COF28022011), which could turn the assignments into some kind of learning objectives for students to achieve. Then, around the middle of the semester when the students had learned most of the speech-making skills, the first assignment could help find out how well students had learned the skills and what aspects needed further improvement. At the same time, through the multiple sources of feedback from peer evaluation, student self-critique, and teacher feedback, students could internalize the features of good speeches and understand their own strengths and weaknesses so that they could strive for better performances in the next assignment. The arrangement of the three speeches allowed the feedback from the first assignment to be fed into the preparation for the second assignment and the same cycle could repeat for the second and third assignments, taking the form of a spiral: preparation 1 → assessment 1 → feedback 1 → preparation 2 → assessment 2 → feedback 2 → ... → feedback 3. In this way, the three assignments could help push students to a higher level in speech making. Therefore, the three prepared-speech tasks were also designed to serve formative purposes.

Furthermore, as mentioned in section 6.2.1, the peer-evaluation and self-critique parts of the prepared-speech tasks were also designed to serve washback purpose.

6.2.2.2 Assessment method

The three prepared-speech assignments belonged to the category of *formal assessment* because they were specified beforehand in the course description (Appendix 21), followed the same procedures for each student, and both Linda and her students regarded them as assessment. Analysis also showed that each of the three assignments involved the following three stages: the preparation stage, the assessment stage, and the follow-up stage.

The preparation stage started from the very beginning of the semester because Linda informed her students of the prepared-speech tasks at the very first lesson when she explained the course description to the students (Linda_COF28022011), but it mainly occurred during the week before the scheduled time for each assignment. In the lesson of week 5, Linda talked about the scheduled time for the upcoming assessment of the informative speech, the use power point slides, and the student sequence of giving their speeches (Linda_COF28032011, Linda_COF31032011). She also sent the marking criteria for the informative speech (Appendix 24) to her students through email. Regarding the two persuasive speeches, Linda also reminded her students of the upcoming assignment one week beforehand, answered her students' questions if any, and sent them the assignment instructions and the marking criteria through email (Linda_EoSI). Moreover, student interviews showed that while they were preparing for their speeches, when they came upon some questions, they could email Linda for advice and Linda usually replied immediately and gave suggestions (e.g., David_EoSI, Lewis_EoSI).

Classroom observation of the week 9 and week 13 lessons revealed that before the first student started his speech, Linda would spend some time preparing her students for peer evaluation. She gave each student a copy of a blank marking sheet (Appendix 24), and told her students whom they were going to do their peer evaluation on and how. She divided her students into three big groups, based on their seating arrangement, with students sitting in one column belonging to one group. Each group of students

were to evaluate the speech of every third speaker. If group one marked the first student, then the next student they would mark was number 4, and the next one number 7, and so on. When one group marked one student, they should also write written feedback for that student. Students from the other two groups should ask questions or give oral comments about the speaker's speech. She also asked the students to pay attention to the aspects and specific points within each aspect in the marking criteria and reminded them to mark conscientiously because the average of their scores on a student would account for 50% of that student's score on that assignment.

After Linda clarified the requirements for peer-evaluation came the assessment stage. Students took turns to go to the front of the classroom and gave their prepared speeches. Classroom observation showed that while one student was delivering his/her speech, the group of students responsible for marking this student took notes and marked on the marking sheet and students from the other two groups prepared slips of paper and wrote comments on them. After a student finished his/her speech, there was always a lot of interaction in class when both the teacher and the students asked questions, gave comments and suggestions, and the speaker responded. The most time that was spent on this type of discussion in my data set was 12 minutes. Those who had written feedback also passed their slips of paper to the speaker after the student returned to his/her seat. Meanwhile, every speaker recorded his/her own speech as well as the follow-up discussion, and the recording served as the basis for them to do self-critique. At the end of a class, Linda always collected students' marking sheets and reminded those who had delivered their speeches to hand in their self-critique journals soon afterwards.

The follow-up stage consisted of students writing up their self-reflection journals and teacher feedback to the individual students. At the end of the semester, both the teacher interview and the student interviews confirmed that Linda gave each student written feedback for each of the three prepared speeches. The feedback contained a mark out of 100 and some written comments. This mark was an average of her score and the average score of all the marks given by the student peers. Linda also gave written comments to students' self-critique journals.

Clearly, each assignment involved the following three stages: the preparing stage during which the task requirements were explained to students repeatedly to drive home the learning goals; the assessment stage during which students took turns to deliver their speeches followed by immediate teacher- and peer feedback; and the follow-up stage during which students reflected on their own performances and the teacher gave written feedback (a score and some comments).

6.2.2.3 Making judgment

Who

As described above, during the assessment of students' prepared speeches, while Linda was a major agent of judgment, the students were also involved in assessing their peers' and their own performances. Therefore, the teacher, students themselves, and their peers were all involved in the assessment tasks.

Construct

Analysis of the marking criteria for this set of tasks (see Appendixes 22 & 23) revealed that these assignments were supposed to assess students' communicative competence, quality of their ideas, and some other aspects (see Table 6.5). For the sake of clarity and brevity, the two sets of criteria were collapsed with those overlapping parts marked with *, those specific to the informative-speech task marked with (I), and those specific to the persuasive-speech tasks marked with (P). The numbers in brackets indicate the frequency of each assessment construct as reflected through the two marking criteria.

Table 6.5 Constructs of Linda's Prepared-speech Assignments as Revealed through the Marking Criteria

Marking criteria	Assessment construct

(P)Topic
    (P)Relevant and appropriate
    (P)Appealing and interesting
	Communicative competence:
	Discourse competence (12)

*Introduction
    *Gained attention
    *Showed relevance of topic to audience
    *Established credibility
    *Introduced topic/thesis statement clearly
    *Previewed body of speech
	Linguistic competence (4)
	
	Paralinguistic competence (5)

*Body
    (P)Structure
        (P)Main points clear and Argument clearly developed
        (P)Demonstrated persuasive organization
        (P)Presented a responsible argument
        (I)Main points clear
(I)Organization effective
        *Language precise, clear, powerful
        *Transitions effective
    (P)Supporting materials
        (P)Strong evidence presented
        (P)Sources fully cited
        (P)Reputable sources incorporated
        *Sufficient number of sources cited
	Ideas:
	Credibility (6)
	
	Relevance (4)
	
	Clarity (2)

*Conclusion (15 pts.)
    *Audience prepared for conclusion
    *Purpose and main points reviewed
    *Closed speech by reference to intro./other devices
	Depth (2)
	
	Truthfulness (1)

*Delivery
    *Maintained eye contact
    *Used voice, diction, & rate for maximum effect
 	
    (P)Used space, movement and gestures for emphasis
    (I)Maintained time limits
    (I)Speak extemporaneously (use note cards/outline only)
	Other:
	Rhetorical effectiveness (9)

*Overall Impression
    (P)Topic challenging
    (I)Topic challenging/interesting
    *Adapted to audience
    (P)Maintained time limits
    *Evidence of preparation & practice
    (P)Was persuasive
    (I)Was informative
	Effort (2)
	
	Ability to meet the time limit (2)

Analysis of the feedback session after each speech in the week 9 and week 13 lessons revealed that both Linda and her students emphasized a lot on the ideas of a speech as well as students' discourse competence and paralinguistic competence (see Table 6.6).

Table 6.6 Aspects Mentioned during the Feedback Sessions of Linda's Prepared-Speech Assignments

Aspect	Example
Discourse competence (28)	
Coherence (13)	You added one point after you talked about trying new things and hard work, you talked about positive thinking. But then in the concluding part, when you are summarizing the main points, you mentioned again trying new things and hard work, but dropping that positive thinking part.
Relevance (9)	I think the target audience should be government officials instead of our students, because these solutions can't be took into actions by our students.
Speech organization (6)	And the structure is actually the Monroe's motivated sequence in a way in my opinion. I think using this motivated sequence is a little bit challenging, but you managed to make us feel that you used the motivated sequence and we are really motivated to some extent, which is really good and effective.
Paralinguistic competence (18)	
Confidence (10)	I think you are more confident and at ease this time compared to your first or former performance.
Eye contact (4)	Well I felt he had made greater effort this time to remain eye contact this time. You remember last time he was, ((The teacher acted the way he presented last time)). always looking down. Yes, he tried to remain eye contact with the audience.
Body language (3)	I think it good for you to add some movements (class laugh loud) like touching your face, but I don't think it necessary for you to touch your face for so many times.
Use of stress (1)	I noticed that you deliberately stressed certain words in your speech maybe in order to let your audience feel to make your main point more clear and make some emphasis on your speech.
Linguistic competence (9)	
Fluency (8)	And you have a good control of your speed of speech, so the audience can easily follow your pace.
Use of parallel structures (1)	I just want to say that she kind of deliberate[ly] used parallel sentence patterns in order to make it [her speech] powerful.
Ideas (24)	
Depth (10)	I really like your two illustrations on the issue, one is the differences between the two cultures and the other is the proper attitudes towards marriage. I think the depth of your illustration is well divided. The first is the reality level, and then the emotional or even moral level. I really appreciate the development of your reasons, and it really deepens our understanding on this issue.
Logic (9)	I still have a question about the experiment you mentioned in your speech: the people losing weight experiment. The researchers supplied people [who] wants to lose weight with calcium supplement, other than ice cream, but your point is eating ice cream will not put on weight?
Clarity (3)	One question I want to ask is that I am not quite clear about your purpose of your speech. Do you mean that you persuade us to do what?
Credibility (2)	Maybe you could have told us that you have recently read about or done some study, I mean establishing your credibility at the beginning of your topic so that , while you didn't do that, that's why I am wondering how did you come up with this topic.
Other (14)	
Rhetorical effectiveness (11)	

Tone (9)	I think it's really a powerful speech, especially because of your tone.
Use of humor (2)	I think the delivery is good, and he uses some humor in the speech to interact with the audience.
Use of visual aids (3)	Last but not the least, this is a suggestion about visual aids. Um I think you should show us the second power point later, not just after the first one.

It can be seen that what was emphasized in classroom matched closely the marking criteria provided.

How

Regarding how judgment was made concerning students' performances during their speeches, clear signs of analytical scoring and criterion-referencing were found, as in Linda's repeated emphasis on the marking criteria (Appendixes 22 & 23) and the students using the marking criteria to evaluate their peers' as well as their own speeches.

In addition, the Q&A session also revealed that sometimes the evaluation was made by comparison with students' previous performances. Not only Linda but some students sometimes made such comments. For example, the following comment was from a student: "I think you are more confident and at ease compared to your first or former performance" (Linda_COF28042011). Therefore, to some extent, the basis of judgment was also student-referenced. There were few signs indicating that Linda compared one student's performance against others', and therefore norm-referencing was not evident.

6.2.2.4 Making use of the judgment

Analysis showed that Linda's evaluation of students' speeches was used in two ways. On the one hand, her evaluation contributed to the final grading of that student. As mentioned in section 6.2.1, each speech weighted 20% in the final grade. Therefore, her judgment was used for reporting purposes.

More importantly, her judgment was used formatively to help with student learning. Classroom observation showed that while she timed each student's speech, she did not time the follow-up discussion after each speech. Though it took much longer time in

her class to finish this task for every student than as scheduled in the course syllabus, she did not shorten the time for follow-up discussion for each speech but just readjusted her follow-up teaching. She explained that she "want(ed) the students to be able to get some immediate feedback from their peers and teacher right after they have made the speeches" (Linda_EoSI). Analysis of the Q&A session after each observed speech showed that a lot of feedback, explanatory in nature, was provided after each speech, pointing out not only the student's strengths but also places for further improvement. This showed that Linda used her judgment to help with student learning.

6.2.2.5 Profile of the prepared-speech assignments

The following figure shows the key features of this assessment practice.
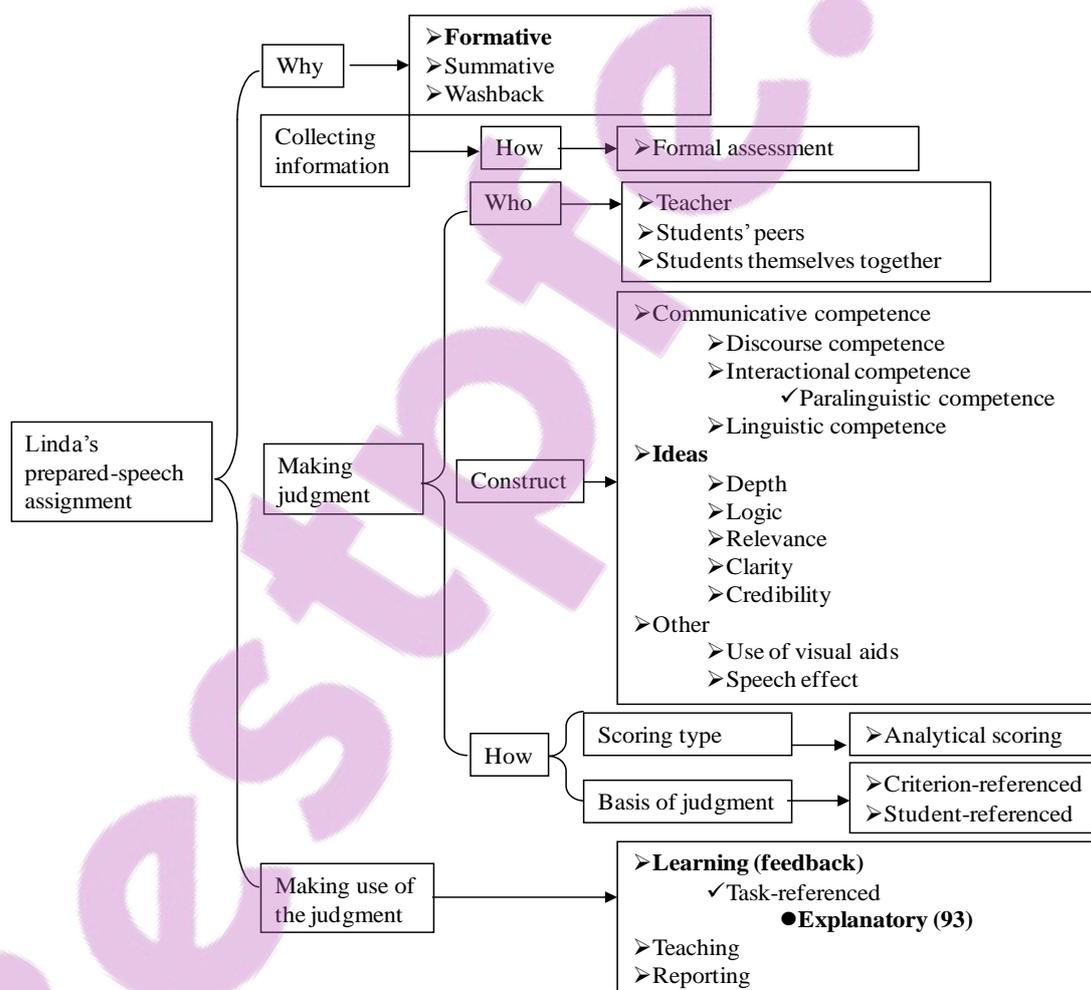
Figure 6.2 Profile of Linda's Prepared-speech Assignments

6.2.2.6 Students' perceptions

Seven students volunteered to participate in the journal writing and student interviews, and their background information can be found in Table 6.7.

Table 6.7 Background Information on the Volunteer Students in Linda's Class

Student name	Gender	English-learning starting age	High school type	Going-abroad experience	Oral English ability
Lewis	Male	9-12	Non-FL	No	Low
Henry	Male	9-12	FL	No	Average
William	Male	Before 8	FL	2 weeks tour abroad	High
David	Male	Before 8	FL	1 month tour abroad	High
Lily	Female	9-12	FL	No	High
Karen	Female	Before 8	Non-FL	1 month tour abroad	Average
Helen	Female	Before 8	FL	No	High

Data from the volunteer students showed that they found this set of assessment tasks very beneficial for their learning and conducive to boosting their confidence.

First, the prepared-speech tasks pushed students to work hard to prepare for the tasks, which helped them internalize the relevant knowledge and skills of public speaking covered during the semester. For example, David, who came from a foreign language school and whose oral English was already at the top level of his class, still spent a lot of time preparing for his speeches. Here is his description of how he prepared for his first persuasive speech:

> I really put a lot of effort into my first persuasive speech. I started previewing the chapters one and a half weeks ago, formed the topic last week and officially started working on it on Saturday night. On that night, I finished the draft. And I refined it over and over again on Sunday. On Monday, I began to prepare the note cards and references. And I spent the remaining two days rehearsing for several times. (David_J3)

For the second persuasive speech, he regarded it as his "final chance to bring out the best of me" (David_J4). Therefore, in addition to preparing for the speech carefully as usual, he also tried out new ideas to organize his speech (David_J4). Clearly, these assignments had pushed him to study hard.

Second, the students found the follow-up discussion/feedback sessions for their own speeches very beneficial and encouraging and helped them to make progress step by step. The immediate and sincere feedback during such sessions not only helped them understand what went well and what went wrong during their speeches, but also boosted their confidence. For instance, Lewis, whose oral English was comparatively poor, felt his first speech "a total failure" because he "was just standing there reading the speech", so "later I began to pay more and more attention to this, I mean, not to read the speech, but to look at your notes and then look at your audience" (Lewis_SRI). For his second speech, he did have some eye contact with the audience (Linda_COF28042011). During the feedback session, while one student pointed out that "your eyes always looking at that direction. It's only you look at your note and then you raise your head and look at that direction," Linda commented in a more positive way: "Well I felt he had made greater effort this time to maintain eye contact this time. You remember last time he was ((Linda acted the way he delivered his speech the previous time)), always looking down. Yes, he tried to remain in eye contact" (Linda_CRT28042011). For his third speech, he talked in a more confident and natural manner (Linda_COF26052011). Although the topic of this speech was not very appropriate, he still felt "I have made some progress this time compared with the previous time. At least I was more confident and paid attention to the eye contact with the audience" (Lewis_J4). It can be seen that informative and encouraging feedback helped him realize his own shortcoming and greatly enhanced his performance and confidence gradually.

Third, the students also found the follow-up discussion/feedback sessions for other students, together with the peer-marking part, helpful for them to understand the qualities of a good speech, which could in turn help them improve their own speeches. Helen, a top student in this class, found that she had learned a lot not only from the teacher's and the classmates' feedback on her own speeches but also from observing others' speeches and evaluating their performances.

> I learned a lot from all of them ranging from their way of delivering to the content of their speeches, to the power and confidence hidden inside their speeches. I tried my best to give evaluation and comments to them on a little piece of paper. Because I think that when commenting on their behaviors I was

also doing some introspective self evaluation, I attempted to learn their advantages and try to avoid the mistakes they made. (Helen_J3)

Finally, those students who did the self-critique part carefully found this task another learning opportunity by reflecting on their own practices. For example, William, a top student from a foreign language middle school, did every self-critique carefully (William_SRI, William_EoSI). Though he thought they were not marked, and there was no word limit, he still did them carefully because he liked the practice that he could listen to his own speech again from the recording, which made him step aside and examine his own performances in a more objective way.

> I found it [self-critique] useful because you know when you look back on the whole preparation and your draft , you know I have recorded my voice when I gave the speech in front of all my classmates, I found it's really helpful for me to reflect on them. … You know it's really good to put yourself in other people's shoes and look at yourself in a more objective way. (Wiliam_SRI)

Therefore, it can be seen that this set of three closely knitted tasks helped improve students' learning and enhance their confidence.

6.2.3 The final exam

The profile of the final exam was generated from the analysis of Linda's textbook, relevant documents (Appendixes 21 & 25), classroom observation field notes relating to the final exam, and the teacher's and the students' end-of-semester interview data. This part follows the same pattern as that for the prepared-speech tasks.

6.2.3.1 Assessment purpose

The final exam was designed for summative purpose only. As Linda believed that speaking skills should be assessed systematically throughout a whole semester, she regarded the final test as another marked exercise, just like the three prepared-speech assignments (Linda_EoSI). She believed that by eliciting a number of students' performances on different occasions, she could get a more comprehensive and more objective picture of a student's public speaking abilities. Besides, classroom observation showed that Linda sometimes asked students to practice giving impromptu speeches in classes (cf. Table 6.9). "Since we have already scored them on three

prepared speeches, it would be justified to see how they perform when they are given a topic to do an impromptu speech" (Linda_EoSI). Therefore, the final test to some extent was to summarize student learning in this course. In addition, the final test was conducted at the very end of the semester, students were not informed of their final test scores, and Linda would not teach them the following semester, all of which strengthened the summative function of this test.

6.2.3.2 Assessment method

The final exam was a very formal assessment practice. According to the observation field notes (Linda_COF16062011), on the day of the test, all the teacher examiners were briefly trained before the test, all the first-year students took the test in the same morning, and every student was asked to deliver an impromptu speech individually to two examiners during the test.

Specifically, on the day of the test, 20 teachers teaching the same grade, though not necessarily teaching this *Public Speaking* course, came as examiners for this final test. They came about 30 minutes earlier than the scheduled time for the test and met at a meeting room where Linda briefed them about the test procedures and the marking criteria (Appendix 25). After the teachers' questions about the test procedures and the marking criteria were clarified, they were paired and went to one of the ten test rooms.

At the scheduled time of the test, students came class by class and waited at a waiting room, where two teachers put them into groups of ten. When it was the scheduled time to start the test, the first group students would each get a piece of paper with a topic written on it. All the students in the same group got the same topic, and every group got a different topic. After the students got their topic, they had five minutes to prepare in the waiting room, during which they could jot down notes on a piece of paper but were not allowed to chat with other students or use dictionaries. When the five minutes was up, they were led to one of the ten test rooms where they delivered their speeches to their examiners.

Observation of Linda's test room showed that after a student came in, the examiners first asked the student to sign his/her name on a piece of paper, and then asked the

student to start the speech. Each speech was timed with a timer and the time limit was set to three minutes. If a student could not finish his/her speech within three minutes, he/she was allowed to talk for 15 more seconds and then was stopped even if he/she still did not finish. Then the examiners asked the student to leave and they discussed about the grading of that student's performance and worked out a score together. There was little interaction between the examiners and the student during a speech.

Because students of the same class came for the test at the same time and they were assigned to ten different test rooms, it was likely that a student might be examined either by his/her own teacher or by teachers from other classes. At the test time, only two of her students took the test at Linda's test room.

After all the students from the eight parallel classes were examined, Linda collected the marking sheets from each test room, and then the *Public Speaking* teachers gathered together and copied down their own students' scores on a different piece of paper.

6.2.3.3 Making judgment

Who

Classroom observation showed that during the test, students were not involved in judgment making in any way. Linda and her colleagues were the agents of judgment.

Construct

Analysis of the marking criteria (Appendix 25) revealed that the final exam was supposed to examine a student's discourse competence, linguistic competence and paralinguistic competence, but it also intended to assess the ideas of a speech and the rhetorical effectiveness (see Table 6.8).

Table 6.8 Constructs of Linda's Final Exam as Revealed through the Marking Criteria

Marking criteria	Assessment construct
Content (30pts)	Ideas
Relevant and appropriate	Relevance

Main points adequately and logically development	Depth and logic
Appealing and interesting	Rhetorical effectiveness
Organization (20pts) Speech has a clear structure (beginning, body, and conclusion) Effective use of signpost words Transitions effective	Discourse competence
Language (30pts) Speaking fluently Language accurate, clear, precise, and powerful	Linguistic competence
Delivery (20pts) Voice loud enough to hear without constraint Maintain eye contact Manipulate voice, diction, and rate effectively Use natural gesture	Paralinguistic competence

Classroom observation revealed that Linda and her partner referred to the marking criteria when marking though they did not give separate scores for each aspect. Her interview data showed that while she looked at the content, organization, and delivery of a speech, she also assessed students' overall language proficiency, because

> I guess it would really show their speaking proficiency, because they really have very little time to think, and within this time they have to come up with ideas, organize ideas, and also think of the language they need to use to express them. So their general language ability can be reflected. (Linda_EoSI)

Therefore, Linda assessed the aspects listed in the marking criteria.

How

The marking criteria (Appendix 25) required the examiners to use analytical scoring since the criteria specified the aspects and weighting for each aspect. However, observation of the final exam (Linda_COF16062011) showed that during the exam, Linda and her partner did not give a separate score for each aspect listed in the marking criteria. Instead, after a student finished his/her speech and left the classroom, Linda and her partner, while referring to the marking criteria, often discussed briefly some striking features of the speech, such as its content, or organization, or language, and then one of them would suggest a score and the other would either totally agree or suggested some changes. Finally they reached an agreement. Because there was usually just one or two minutes for the two examiners to exchange ideas and reach an agreement, their discussion was usually very brief. This indicated that Linda practiced

holistic scoring to some extent, though she might have considered those aspects listed in the marking criteria.

Regarding the basis of judgment, the final test was primarily criterion-referenced, as reflected through the marking criteria, but observation revealed that there was also a norm-referenced element. For the first two or three students, Linda and her partner did not put down their holistic marks immediately. Instead, they waited until they examined more students and compared these students' performances so that the scores would indicate the rough ranks of these students (Linda_COF16062011). Clearly, both criterion-referenced scoring and norm-referenced scoring were involved in their marking.

In addition, as mentioned in section 6.2.3.1, in Linda's teaching context, the final test was no longer an individual teacher's responsibility, but the whole teaching group's responsibility. A certain level of standardization had been achieved through brief teacher training, arrangement of the students, and the consistency concerning the preparation time and test time for each student. These measures could all enhance the quality of the final test, which should be regarded as moderation measures (Harlen, 2007). Since many teachers were involved in the test, such quality assurance measures should help enhance the reliability of the marking because they might reduce some random sources of measurement error (Bachman, 1990).

6.2.3.4 Making use of the judgment

Linda reported that the test scores were used to generate the final composite scores for the students and she did not have another chance to provide any feedback to her students on their test performances (Linda_EoSI). Therefore, her judgment was solely used for the reporting purpose.

6.2.3.5 Profile of the final test

The following figure summarizes the key features of the final test.

Why → ➢Summative

Collecting information → How → ➢Formal assessment

Linda's final test

Making judgment → Who → ➢Teacher

Construct →
➢Communicative competence
　➢Discourse competence
　➢Interactional competence
　　✓Paralinguistic competence
　➢Linguistic competence
➢Ideas
　➢Relevance
　➢Depth
　➢Logic
➢Other
　➢Rhetorical effectiveness

How →
Scoring type → ➢Analytical + holistic scoring
Basis of judgment → ➢Criterion-referenced ➢Norm-referenced
Moderation → ➢Standardization in test administration

Making use of the judgment → ➢Reporting

Figure 6.3 Profile of Linda's Final Test

6.2.3.6 Students' perceptions

Though the final test took the form of an impromptu speech, all the volunteer students prepared for it. They all found some possible topics from the Internet and practiced by themselves. For example, David forced himself to "complete a speech within three minutes" (David_EoSI), and Lewis tried to "come up with a few key points within the limited time" (Lewis_EoSI). Lily practiced hard during the two days before the test to "find the feeling of doing this kind of speech" (Lily_EoSI). William reviewed the chapter on impromptu speech in the textbook and practiced the topics given in the textbook (Wiliam_EoSI).

However, all the volunteer students reported in their end-of-semester interviews that they were not very much worried about their final test results. As Lily said, "everybody should get a fairly good score" because "we have already got a large

percentage of the final composite score based on our daily classroom performances and the final test score only accounted for a small percentage" (Lily_EoSI). Therefore, some of the students regarded the test as another marked exercise and did not feel very much pressured (Wiliam_EoSI, David_EoSI, Lily_EoSI, Henry_EoSI).

6.2.4 Unrecognized CA

Linda's idea of assessment was generally restricted to those activities that were marked and contributed to the final grading (Linda_BI, Linda_EoSI). However, analysis of her daily classroom teaching revealed that some of her classroom activities could also be viewed as CA practices. This section will focus on such classroom embedded CA practices.

6.2.4.1 Data

The profile of Linda's unrecognized CA practices was generated from the detailed analysis of the classroom recording transcripts of the weeks 1, 5, 9 and 13 lessons and the relevant data from participants' interviews and student journals. Thursday's lessons of week 9 and week 13 were devoted to the prepared-speech tasks and therefore were not included in the following table. Table 6.9 provides a brief description of the classroom activities during the observed lessons.

Table 6.9 Linda's Classroom Activities

Duration in class	Classroom activity	Brief description
Week 1		
Thursday		
10:13-10:42	Student giving a prepared speech	Topic: The value of public speaking Source: Homework from last lesson Format: Students preparing individually → Students working in pairs → Two students delivering their speeches to the whole class → Feedback from Linda and other students after each speech
10:43-10:51	a) Q&A	Topic: Similarities and differences between public speaking and daily conversation Source: Homework from last lesson (students were asked to read chapter 1 of their textbook on public speaking) Format: Linda raised the question "What are the similarities and differences between public speaking

		and daily conversation" to the whole class → individual students volunteered answers and Linda provided feedback after each response → Linda showed her power point slide which contained the main points she had prepared
10:52-11:05	Q&A	Topic: How to handle nervousness when giving a speech Source: Homework from last lesson (students were asked to read chapter 2 of their textbook on public speaking) Format: The same as that for activity a)
11:06-11:09	Break	
11:10-11:24	b) Teacher-guided appreciation of a sample speech	Source: A sample speech from chapter 3 of the textbook Format: Students watched the speech → Linda conducted a class discussion on the good qualities of the speech through many IRF sequences
11:25-11:40	c) Teacher lecturing	Topics: Plagiarism & Different types of listening Format: Linda talked through many power point slides on the two topics and there was little teacher-student interaction
11:40-12:00	A sentence-matching game	Format: Each student is given one slip of paper with a sentence on it. Each sentence is matched with another sentence. Students should find their match and explain the meaning of their sentences.
Week 5		
Monday		
9:00-9:10	d) Student news report	Format: One student gave a news report in front of the whole class with the help of ppt. No feedback from Linda or other students.
9:11-9:15	Teacher's reminder of the first prepared speech assignment	
9:16-9:38	Student making a prepared speech	Source: Homework from last lesson (students were asked to revise their previous week's speeches) Format: Students worked in groups of four, with two students giving their previous week's speeches again and the other two students as listeners and giving comments afterwards → Linda asked one listener from each group to comment on the improvements their group members have made
9:39-9:45	Q&A	Topic: How to make an informative speech Source: Homework from last lesson (students were asked to read chapter 13 of their textbook on public speaking) Format: The same as that for activity a)
9:46-9:55	Teacher-guided appreciation of a sample speech	Source: A sample speech from chapter 13 of the textbook Format: The same as that for activity b)
Thursday		
10:17-10:25	Student news report	Format: The same as that for activity d)
10:26-10:29	Teacher's reminder of the first prepared speech assignment	
10:30-10:48	Teacher lecturing	Topic: How to gather materials for a speech

211

		Format: The same as that for activity c)
10:49-11:12	Student conducting an interview	Topic: Attitudes towards fake and shoddy goods Purpose: Practice conducting an interview, one way of gathering materials for a speech, which Linda mentioned in her lecture just now Format: Students worked in groups of three and interviewed their group members on their attitudes towards fake and shoddy goods, while Linda moved around and observed.
11:13-11:22	Break	
11:23-12:00	Student making an impromptu speech	Task: Students were asked to make an impromptu speech based on the interview data they collected. Format: Students preparing individually → Students working in pairs → Three students delivering their speeches to the whole class →Feedback from Linda after each speech
Monday of Week 9		
9:05-9:09	Student news report	Format: The same as that for activity d)
9:10-9:23	Q&A	Topics: Logical fallacy & How to appeal to emotions Source: Homework from last lesson (students were asked to read chapter 14 of their textbook on public speaking) Format: The same as that for activity a)
9:24-9:37	Teacher-guided appreciation of a sample speech	Source: A sample speech from chapter 14 of the textbook Format: The same as that for activity b)
9:38-9:57	Student making an impromptu speech	Topic: Should Yao Jiaxin (a criminal) be sentenced to death? Format: Students preparing individually → Students working in pairs and Linda observing one group → Teacher-led discussion on the points students had come up with for their speeches, how to make each point clear and well-developed, and time management.
Monday of Week 13		
9:00-9:06	Student news report	Format: The same as that for activity d)
9:07-9:24	Teacher lecturing	Topics: How to make a power point presentation Format: The same as that for activity c)
9:25-9:50	Student making an impromptu speech	Topic: Does appearance matter? Format: Students preparing individually → Students working in pairs → Three students delivering their speeches to the whole class →Feedback from Linda and other students after each speech

It can be seen from the above table that Linda' classroom activities can be grouped into five major types (the first five types in Table 6.10). Further analysis revealed that three kinds of her classroom activities could be regarded as *(*activities marked with a * in Table 6.10). Though "student news report" was a regular activity at the beginning of each lesson and there was requirement for it in the course description (Appendix 21), this activity was not regarded as a CA-involving practice because there was a lack of

evidence for the third step of assessment. Classroom observation revealed that usually after a news report, Linda would directly move to the lesson rather than spending some time commenting on the student's presentation. Linda's explanation showed that this task was mainly designed as a practice opportunity rather than as an assessment activity.

> The purpose of this task is just to draw students' attention to the current affairs, cos we have noticed from previous years that some of the students, particularly girls, they just don't mind what's happening around them and in the world, and we think it's not good. So, uh, and also presentation has a lot of things similar to public speaking, so it's an opportunity for them to really practice that. And also like if you have noticed in the past few weeks, there are just a few students who had the chance talking in front of the whole class. This is another opportunity for having more students to practice speaking in front of the whole class. (Linda_SRI)

Table 6.10 Overview of Linda's Classroom Activities

Classroom activity	Frequency of occurrence	Percentage of total observed class time
*Student making a speech	5	37%
Teacher lecturing	3	14%
*Q&A	4	11%
*Teacher-guided appreciation of a sample speech	3	10%
Student news report	4	8%
Student conducting an interview	1	6%
Game	1	6%

It can be seen that similar to Andrew's and Mary's, more than half of Linda's class time was engaged in assessment-involving activities. In what follows, such CA practices will be described in detail.

6.2.4.2 Unrecognized CA practices

It should be pointed out that this group of assessment-involving activities shared the following two features in common. First, they all served formative purpose, as can be seen in the following description. Second, since Linda did not consider such activities as assessment, such categories as *scoring type*, *basis of judgment, and moderation* were found mostly irrelevant, except for the speech-making activities which showed some evidence of Linda's basis of judgment. Therefore, to avoid repetition, the

213

description of this group of unrecognized CA practices will focus on the information-collecting method, the construct component of the making-judgment step, and the making-use-of-the-judgment step. Students' perceptions of each type of activity will be presented after the profile of that type.

CA practices within Q&A sessions

As can be seen from table 6.10, Linda adopted two major ways to deal with the key points concerning public speaking skills: either through teacher-led Q&A sessions or through her own lecturing. Linda explained that she conducted Q&A sessions on the key points mentioned in the textbook, because she wanted to "check if they [the students] had read the chapter and what they had got from the chapter, and also to find out if there was anything important they had overlooked" (Linda_SRI). However, sometimes Linda gave detailed explanations herself rather than engaging the students in Q&A interactions, such as when she gave a lecture on how to gather materials for a speech at the week 5 lesson, because "this chapter, Gathering Materials, is not in the book. It's something else. It's from another book" (Linda_SRI). It can be seen that the Q&A sessions should be regarded as planned assessment-involving instructional activities because they were used to find out if students had read the textbook and to what extent they had understood the required chapters, as can be seen from excerpt 6.1.

This excerpt is taken from the Monday lesson of week 5 when Linda was talking about the strategies for making an informative speech, key points in Chapter 13 of their textbook.

Excerpt 6.1

Turn	Speaker	Transcript	Feedback type
1	T	[...] Now let's spend a little time checking on your understanding of the main points from the chapter Speaking to Inform, Chapter 13. Now very quickly. What are the four types of informative speech discussed in the chapter? What are they? Four types? Remember? Remember?	
2	SS	Objects.	
3	T	The first is objects. Objects can include—Ok, first of all, objects. And then?	Confirmative

9	T	[...] ((Students got all the four types correct.))	
9	T	OK. Then, why must the speaker not to overestimate what the audience already knows about the topic and what can you do to make sure that your ideas don't pass over the heads of your listeners?	
10	SS	((No student responded. Some students looked through their textbooks.))	
11*	T	Then what are the suggestions given in the book to prevent this from happening, that is, passing over the heads of the listeners? What can you do?.... Uh? ... what? ((She observes the class and expects some volunteers to answer this questions.))	
12	S1	We can use examples.	
13	T	OK. By using examples to illustrate. Any other means?	Confirmative
		[...] ((Three other students provided their answers and Linda acknowledged their answers.))	
20	S5	When you put forward a technical term, you can give your listeners some explanation and if necessary, you can use your body language to show the meaning.	
21	T	Exactly, yes, remember to explain, particularly when you are preparing, you can anticipate that probably at this point I maybe need to put in some explanation. This is important. Don't assume that it's so easy, everyone should know, but maybe not.	Confirmative Explanatory

In this excerpt, Linda elicited students' knowledge about informative speech through questioning (turns 1 & 9). Both questions were about the key points of chapter 13 of the textbook and both were prepared in advance because they were on her power point slide. When a student's answer was correct (turns 2 & 12), she usually provided confirmative feedback (turns 3 & 13). When she noticed that students had difficulty (turn 10), she usually rephrased her question and elicited again (turn 11), which was a kind of scaffolding in nature. The episode (turns 11-13) is regarded as an incidental assessment episode because Linda's question (turn 11) was contingent upon students' performances (turn 10) and not planned beforehand. When a student provided a good answer (turn 20), she not only confirmed it but also explained why it was a good answer (turn 21).

Further analysis revealed that the Q&A sessions essentially consisted of the traditional IRF sequence (Sinclair & Coulthard, 1975) or its variations. For example, sometimes, Linda would allow several students to provide their answers before she gave her feedback (Linda_COF03032011), making an IRR(R)F pattern. While most of such segments should be regarded as planned assessment-involving instructional episodes, because Linda's questions were prepared beforehand, ten incidental assessment

episodes were also identified, some of which were rephrases of a previous question (e.g., turn 11 in Excerpt 6.1) while others were better versions of students' responses, as reflected through the following example.

During the task on how to handle nervousness in the week 1 lesson, Linda asked students what they were afraid of, and one student said:

S: When I standing on the stage, I just don't know what to say.
T: Ok, so you are afraid of being on the stage. Ok, stage fright.
S: Yes.

As Linda pointed out, this group of Q&A activities were to check students' knowledge about the key points contained in the textbook (Linda_SRI). Her judgment of students' responses to her questions was used mainly to provide feedback to consolidate their understanding of those key points. Her feedback was primarily confirmative (25 instances) and explanatory (16 instances). Very often she would first confirm and then provide an explanation (see turn 21 in Excerpt 6.1 for an example).

During such Q&A sessions, Linda was always the questioner and students were not involved in self-assessment or peer-assessment.

Student interview data showed that some students found such sessions useful while others found them unnecessary. For example, Lily found Linda's Q&A sessions on the key information from the textbook a "good reminder of the key points" (Lily_SRI), and Wiliam felt such sessions "strengthened [out] his understanding" of the textbook (William_SRI). However, Henry thought that what Linda emphasized was already "in the textbook," "and it's too obvious and anyone who answers the questions, all of the students will knew [know] that she or he is reading the book. So we prefer not to answer the question" (Henry_SRI).

CA practices within speech-appreciation activities

Classroom observation showed that Linda sometimes guided the class to evaluate a sample speech from the textbook. During such activities, she usually drew students' attention to a couple of questions shown on her power point before she played the video. After playing the video once or twice, she usually conducted a class discussion,

guiding the students to evaluate the speech using the principles they had learned. Based on what students said, she usually provided immediate feedback or raised further questions to help students see the good points of the sample speech. It can be seen that such speech-appreciation activities involved assessment and should be regarded as *prepared assessment-involving instructional activities*.

Linda explained that because the speech "is kind of like a demonstration, it's a model speech," through such speech-appreciation activities, she wanted the students "to really see the strength of it" (Linda_SRI). It can be seen that the purpose of such activities was *formative*, aiming to consolidate the principles and strategies of giving an effective speech she had mentioned in class.

Analysis revealed that the class discussion period contained both *planned assessment-involving instructional episodes* and *incidental assessment episodes*, as can be seen in excerpt 6.2. This excerpt is taken from the Thursday lesson of week 1 when Linda guided the students to evaluate a sample speech. On her power point slide, she showed the following three questions before she played the speech to the whole class.

What are the points made in the speech?
How does she start and end the speech?
Do you think it is an effective speech? Why?

After playing the video twice, she conducted a class discussion. The following excerpt is taken from this discussion period.

Excerpt 6.2

Turn	Speaker	Transcript	Feedback type
1	T	What do you think? What do you think she is talking about? Did you get it?	
2	SS	Yes. The Olympics.	
3*	T	Yes, this is the bidding of the 2008 Olympics. But then what is the theme of her speech? Did you get that? Yeah, what is she talking about?	Confirmative
4	SS	Cultural aspects.	
5	T	Cultural programs. Yes, she is introducing cultural programs of China, yes. Um, yeah, anything to say after you've watched it? Anything you've noticed?	Corrective
6	S1	She is neither arrogant nor humble. She just confidently	

		speaks.	
7	T	Yes yes, she did it very confidently. Then it gives you the feeling that she speaks very naturally. Right? Not like you know some people was thinking 'Oh, I'm making a public speech, so I need to' ((The teacher pretends to be serious.)) You know ((laugh)), but she is very natural, yeah. And, any other things? ((Linda indicates another student to talk.))	Confirmative Explanatory
…	…	[...] ((Three other students made some comments about the structure of the speech, the speaker's eye contact and attitude, and Linda confirmed their ideas after each comment.))	
14	S5	And she is humorous. She told jokes about Chinese football.	
15*	T	Yes, yes. I want to say a little bit more on that. Remember I heard some response from you when she said 'now you can understand why our women's football team does so well'. Yeah, I heard some response from you, which reflects what you were thinking on hearing that. Can you explain to me what that noise you made means? Were you impressed by it or what?	Confirmative
16	S6	Maybe the women's football team was not playing that well, but the renovation is interesting.	
17	S7	Well, I think now the women's team is not so good, but in the year of 2001, it's really a stronger team.	
18	S8	[...] ((Another student gave a similar opinion.))	
19	T	Um. Yeah. Right. Viewing the speech now, because the women's football team is not doing really well now, we get different responses from the audience. The thing I want to say about this, ok, can be a technique in a speech. That is, to relate to the audience by using a specific example and also relate this example to what she said previously. But then you've got to do it well. It's a little bit like using humor in your speech. If it is done very well, good. It has a positive effect. But if it is awkwardly done, then it would have opposite effect. So you've got to be careful. If it can be done well, good. If not, if you are not sure, maybe it's a good idea not taking the risk.	Confirmative Explanatory

It can be seen that this excerpt contained several assessment episodes. In the first segment (turns 1-5), Linda asked a question to elicit students' comprehension of the speech (turn 1). When she noticed that the students failed to grasp her question (turn 2), Linda rephrased her question to scaffold the students (turn 3). When she found some students provided a relevant though not quite accurate answer (turn 4), she gave a more precise expression as a kind of corrective feedback (turn 5). Clearly, Linda's re-elicitation and feedback were based on the assessment of her students' actual performances, and therefore turns 3-5 is an incidental assessment episode.

218

In the second segment (turns 5-7), Linda asked a very general question to elicit students' evaluation of the speech (turn 5). Then she noticed that one student's idea was to the point (turn 6) and therefore provided confirmative feedback and even elaborated on it (turn 7). This episode (turns 5-7) is regarded as a planned assessment-involving instructional episode, because Linda had clearly stated the purpose of this task was to evaluate the videoed speech and her general questions like "anything to say after you've watched it" and "any other things" were not based on students' actual performances but were to invite students to mention some aspects other than those already mentioned. Therefore, similarly, the other three students' opinions and Linda's respective feedback are all regarded as planned assessment-involving instructional episodes.

Actually, Linda's general question (turn 5) also covered the third FA episode (turns 14-19) presented in excerpt 6.2. When she noticed that one student touched upon an important aspect (turn 14), Linda prompted the class to reflect on the issue (turn 15). It was like a scaffolding to push students to reflect further on this aspect. After evaluating three students' opinions (turns 16-18), Linda gave confirmative feedback and further explanation on it, pointing out the importance of relating to the audience (turn 19). Clearly, in this episode, Linda's scaffolding question (turn 15) and confirmative and explanatory feedback (turn 19) were based on her assessment of her students' actual performances and should be regarded as two incidental assessment episodes.

Analysis of all the three speech-appreciation activities in the dataset revealed nine planned assessment-involving instructional episodes and four incidental assessment episodes. It can be seen that through such discussions, Linda could foster students' abilities to use the knowledge and skills they had learned about public speaking to evaluate a speech. In other words, this group of activities focused on students' abilities to evaluative a speech.

During such speech-appreciation activities, Linda had always been the questioner or the guide and students were not involved in evaluating each other's responses. Her

feedback was primarily confirmative (13 instances), though there were also four instances of explanatory feedback and one instance of corrective feedback.

Some students found that Linda's guided analysis of a videoed speech helpful for their learning. For example, when asked about the most impressive part of the lesson of week 1, Lily wrote "the video of Yang's speech left me deep impression," because the speech "was informative and opened the door of Chinese culture to people around the world", Yang "performed well on stage and showed a welcoming attitude", and Yang "involved some humour, such as the Chinese football team" (Lily_J1). Since the discussion session on this speech touched upon such aspects as "cultural programs," "Yang's manner," and "the use of humour" (cf. Excerpt 6.2), it should be proper to say that Lily's learning was somewhat influenced by those FA episodes in this activity.

CA practices within student speech-making activities

Classroom observation showed that Linda spent comparatively more class time on student speech-making activities (cf. Table 6.10). While she sometimes asked students to give a prepared speech and sometimes an impromptu speech, the way she conducted these activities, the fact that she always provided feedback on some students' speeches and sometimes even encouraged students to give peer feedback (cf. Table 6.9) indicated that assessment was involved (an example can be found in excerpt 6.3). Since these activities were all prepared in advance because the instructions for these activities were all in her power point slides, these activities should be regarded as planned assessment-involving instructional activities.

Analysis showed that the purpose of her carrying out these activities was formative in nature. In her Monday lesson of week 5, Linda asked students to give their previous week's speech again. She explained that

> It's kind of like giving a second draft of writing. Remember last time they did their speech, and then I asked them to evaluate their central idea, main points, and everything, so I asked them to go back and revise their speech based on the comments they've got. Yeah, and then we can see whether they've really got it or not, and how much they can improve. (Linda_SRI)

It can be seen that this activity was essentially a homework-checking activity, through which Linda intended to assess if her students had taken in the feedback they had got during the previous lesson and if they had made any improvement. In her Thursday lesson of week 5, Linda asked students to make an impromptu speech based on the interview data they just collected. Linda explained the purpose behind this activity in the following way:

> One is to see if they can rightly summarize the main points they've got from their interviewees, like the information they have gathered for their speech, then uh, because this is just in class time, time is really short, and I don't want them to give another prepared speech, and also I think the ability to give impromptu speech should also be practiced. So this is a time for them to do that. (Linda_SRI)

It can be seen that Linda intended this activity to be a practice opportunity for students to practice summarizing information for a speech as well as to make an impromptu speech. Since Linda was mainly concerned with student learning rather than grading students, these activities were for formative purposes only.

Analysis of all the speech-making activities revealed that while each of these activities was a planned assessment-involving instructional activity, within them there were also eight incidental assessment episodes where Linda provided scaffolding questions to guide the students to see their own strengths and weaknesses and find their way to improve, as can be seen from excerpt 6.3.

This excerpt is taken from the Thursday lesson of week 5. Linda had asked the students to conduct interviews in groups on people's attitudes towards fake and shoddy goods. After this group work, Linda asked the students to make an impromptu speech based on the interview data they had just obtained. Students practiced their impromptu speeches first in groups before Linda asked three students to deliver their speeches again to the whole class. Excerpt 6.3 shows the first student's impromptu speech and Linda's feedback on it, which reflected her assessment of this student's speech.

Excerpt 6.3

Turn	Speaker	Transcript	Feedback

			type
1	S1	Just now I have made an interview with my desk mates, and we talked about something about fake goods. And we discussed and my interviewee thought that there are some kind of fake goods that are acceptable and practical, but some of the fake goods are just is better for us not to buy something fake goods. And my interviewee thought that for something like clothes and shoes that are easily wear off and of good quality. And it is :: economical and practical to buy something some fake goods like that. He shared with me one of his experience last weekend. [...] ((He told that example here.)) But for some of the daily use, for some food, food and drinks, my interviewee thought it was not so sensible to buy some fake ones. ((He told another example.)) So we'd better be observant and to check when we buy some daily products. That's all about the whole interview. ((The class applaud.))	
2*	T	Um, that's a report of the answers you got from your interviewee, right? If we are going to summarize, can we summarize from the answers that the kind of, say, can we summarize, never mind the number of interviewees, cos we are doing it in class and you only had two, so never mind that. Can we summarize in one sentence the attitude of your interviewee towards fake goods from the answers?	Explanatory
3	S1	Yes.	
4*	T	Like what? If you are asked to summarize in one sentence.	
5	S1	Maybe not all fake goods should be rejected and some of them are practical.	
6	T	Ok, ok, then that can be one of your points if you are making a speech, right? So that's why we do research. We get a lot of information, answers from people, but then we need to highly summarize, Ok. Ok. ((She indicates another student to give her speech to the class.))	Explanatory

In this excerpt, after one student delivered his speech (turn 1), Linda noticed that what the student did was not an impromptu speech but a report of the interviewee's answers to his questions. Then she recognized "a mismatch between my expectation and how they understood what I wanted them to do" (Linda_SRI) (turn 2). Instead of telling the student directly he was wrong, Linda asked the student to synthesize the main idea of his interview data in one sentence (turns 2 & 4). When she recognized that the student was able to do it (turn 5), she further pointed out that this summary sentence could serve as one key point in his speech (turn 6), an explanatory type of feedback telling the student how to improve.

The importance of synthesizing the interviewees' ideas in the speech was re-emphasized after the second student delivered her speech, because this student

made a similar mistake. Probably due to Linda's repeated emphasis, the third student did a better job, and Linda gave more positive comments afterwards.

> Um we can see some summary work done in this report, uh? It's not like this is the question this is the answer. She did some summary job. And also it was good that before she ended the speech, she kind of restated it. ... a conclusive remark is that my interviewee would refuse to buy for two reasons. That was also quite clear. Good! (Linda_CRT31032011).

Clearly, assessment was involved in this episode, as reflected through the adjustment she made based on students' performances. This was confirmed by Linda's interview data as well.

> I was in one group, and this girl was giving a report. Actually it was not what I was expecting. They were just reporting, this is the question and this is the answer, this is the second question, and this is the answer. It was not a speech at all. Then I asked one boy and another girl to give their speeches after this group work, and they all did the same thing. So I found it was not a speech at all. So I kind of changed, not really changed, I wouldn't want to tell them directly, 'no, it was not right,' I just maybe as a chance to point out to them like ok those are the information, and then how to make the information and put them into the form of a speech. It's like this process. Like I was expecting that kind of product, but they were somewhere else. There is like a step in a process. So I realize maybe I should move back to that stage in the process instead of expecting an end product. (Linda_SRI)

Classroom observation revealed that while Linda had always been the major agent of judgment during these speech-making activities, sometimes she also invited students to give peer feedback before she gave her own comments (Linda_COF03032011 & Linda_COF23052011). Linda believed that "Learning to judge other people's work is one important step in learning how to make a speech" and "Asking them to comment on each other's work could also give them a feeling that the class is a learning community in which they help each other" (Linda_EoSI).

For the speech-making activities, analysis of both the peer feedback and Linda's feedback revealed that Linda and her students emphasized a lot the ideas of students' speeches, in addition to their discourse competence as reflected through the organization of their speech and their paralinguistic skills as reflected through their non-linguistic delivery skills (see Table 6.11).

Table 6.11 Aspects Mentioned during the Feedback Sessions of Linda's Speech-making Activities

Aspect	Example
Ideas (24)	
Credibility (10)	...he needs some more examples to support his idea, like the last he talked about something about experiences... anyway, he may need some more examples.
Depth (4)	... that's a report of the answers you got from your interviewee, right? If we are gonna summarize, can we summarize from the answers ...
Relevance (4)	It's that you didn't really directly address the topic, the value. Did you talk about the value of public speaking? It was just, you are just mainly sharing your own experience, right?
Logic (3)	Yes, yes, people argue that yeah you cannot choose the face in fact, but it doesn't mean that it doesn't matter. Yes, yes, because some people, well, I am not born pretty, I am not born handsome, but I can go have plastic surgery, yes, if I really care, if I really want to. Ok.
Clarity (3)	I think he clears some of his ideas this time.
Discourse competence (10)	
Speech organization (10)	I think the structure is better than last time.
Paralinguistic competence (7)	
Voice quality(3)	...his voice is loud enough for me to hear because I sit in the farthest position from him.
Body movement (2)	I noticed that in the beginning of his speech, he folded his hand like this ((the student crossed her arms)) and later on he unfold his arms and took his hand downward. So I think that's good because maybe he was conscious about his behavior and wanted to appear more open.
Confidence in speaking (2)	... she overcomes her stage fright. Although she said she was still shaking, I could not see it at all.

A comparative analysis revealed a high level of consistency between what was emphasized during such speech-making activities and the marking criteria for the prepared-speech tasks (cf. Table 6.6), the marking criteria for the final test (cf. Table 6.9), those aspects actually emphasized during the prepared-speech tasks (cf. Table 6.8) and the final test, and what was emphasized in the textbook. They all focused on the ideas of a speech, the organization of a speech, and the speaker's non-linguistic delivery skills. The only difference lay in the relative emphasis on linguistic competence. While both the prepared-speech tasks and the final test included students' linguistic competence as one aspect to be measured, this aspect was not emphasized much during the speech-making activities embedded in classroom teaching.

This consistency indicated that during such speech-making activities, Linda's judgment was criterion-referenced, against the key principles or strategies suggested in

the textbook, which were also integrated into the marking criteria for the prepared-speech tasks and the final test. When she sometimes encouraged students to provide peer feedback during some speech-making activities, Linda also emphasized that students should "establish some criteria" in their own minds and "evaluate a speech according to these criteria" (Linda_EoSI). It can be seen that Linda strove to promote student self-regulated learning by equipping them with the evaluation tools which could serve both as learning goals and criteria for making judgment.

Besides, Linda also tried to help students see their own progress and therefore her judgment was sometimes also student-referenced. Linda believed that it should be more important to help the students see their own progress than giving them a score against some fixed criteria.

> I think what I care most is if they learn, I don't care like kind of scores they get, because each person might start from different level. So according to one criteria, say for one student who was not that proficient, maybe this person has made a lot of progress. But if you use that criteria to judge this person, it would still be like an average score. But actually to this person, it has been a lot of development, a lot of achievement already. And I don't want them to feel discouraged when they see their scores. So every time I really don't want them to feel like that. So that's why I don't like to score students. (Linda_EoSI)

Classroom observation showed that sometimes Linda specifically asked the students to comment on the progress their classmates had made. For example, in the Monday's lesson of week 5, she asked the students to give the speech a second time and asked those students who had listened to the speech twice to specifically comment on the speech so as to help the speaker to see the progress.

Analysis revealed that Linda's judgment during such speech-making activities was mainly used to provide feedback to help with student learning. For example, in the Thursday's lesson of week 1, after one student gave a speech and some students commented on this speech, Linda first acknowledged that the speaker's personal experience was attractive but also pointed out that the speech did not really address the given topic, and she added that if students did not have an interesting personal experience, they could use anecdotes of other people (Linda_COF03032011). Thus, Linda's judgment was used to provide feedback to let students know what they did

well or not well and inform students of the places for further improvement. Her feedback was all explanatory in nature, 41 instances in total. In addition, her judgment sometimes was also used for modifying her teaching (cf. Excerpt 6.3 and its following discussion).

In comparison with the Q&A sessions and the speech-appreciation activities, more students mentioned the usefulness of the speech-making activities for their learning. While there were only six instances commenting on the above two types of FA from the collected data, there were 17 instances on the speech-making activities. Analysis of the students' comments showed that such activities not only helped the students grasp some speech-making strategies, but also enhanced their critical thinking ability.

Most of the students' comments were about their realization of some speech-making strategies. For example, after the Thursday lesson of week 1, Henry remembered

> one of our classmates pointed out that William folded his arms at the outset of his speech which was beyond my observation. This reminded me of once learning that 'folding arms' would send 'unwelcome' messages to others so I bear in mind that such gestures must be excluded in my future speeches. (Henry_J1).

For the activity of asking the students to give a speech the second time during the Monday lesson of week 5, Karen felt "reviewing and restating the speech is a very necessary and effective way for improving the public speaking skills. It's more useful than simply stating the skills mentioned in the textbook" (Karen_J2). For the impromptu-speech activity of week 13, Lewis felt that after observing his classmates' speeches, he should add specific examples in his speech; otherwise his speech would be too dull (Lewis_SRI); William had learned to limit his main points to two so that he could elaborate on each (Wiliam_J4); and Lily had realized that "I should slow down when I make a speech or I would fail to explain logically and become nervous" (Lily_J4).

Moreover, some good students felt the impromptu-speech activities were a good way to train their critical thinking. For example, Helen considered impromptu speech a very "challenging" but also very "important" type of speech and therefore hoped "this class could provide us with more opportunities to make impromptu speeches"

(Helen_J3). Near the end of the semester, she realized that such activities were very useful in improving her critical thinking ability.

> The practice of making impromptu speeches mainly helps us better organize our thoughts and words in a very short period of time, and also helps us to try to think things from different perspectives as much and thorough as possible. I find that I could be more quick-witted than before, but I think I still need to practice more by myself after class in order to be more capable to do a better job. (Helen_J4)

In addition, Henry particularly liked the format of asking some students to give their speeches before the class and then inviting the class to comment on the speeches, because in this way, he could "listen to others' speeches" and then "compare theirs with my own so as to "find the advantages and also shortcomings that require improvement" (Henry_J1).

In spite of such positive comments on Linda's FA practices, in the end-of-semester interviews the volunteer students felt such practices were not as helpful for their learning as the three prepared speeches, because they were "just classroom activities," "not that formal", "not graded" (Helen_EoSI, William_EoSI), and the students "did not have to prepare for them" (Henry_EoSI). In contrast, they all thought the three prepared speeches most helpful owing to the whole process from "preparing for it," including "doing research, writing up the speech, and rehearsing again and again" (Henry_EoSI), to the "inspiring and encouraging feedback" after delivering the speech (David_EoSI), to finally reflecting on the quality and the delivery again through self-critique and Linda's feedback (William_EoSI).

6.2.4.3 Profile of unrecognized CA practices

The following figure summarizes the key features of Linda's unrecognized CA practices.
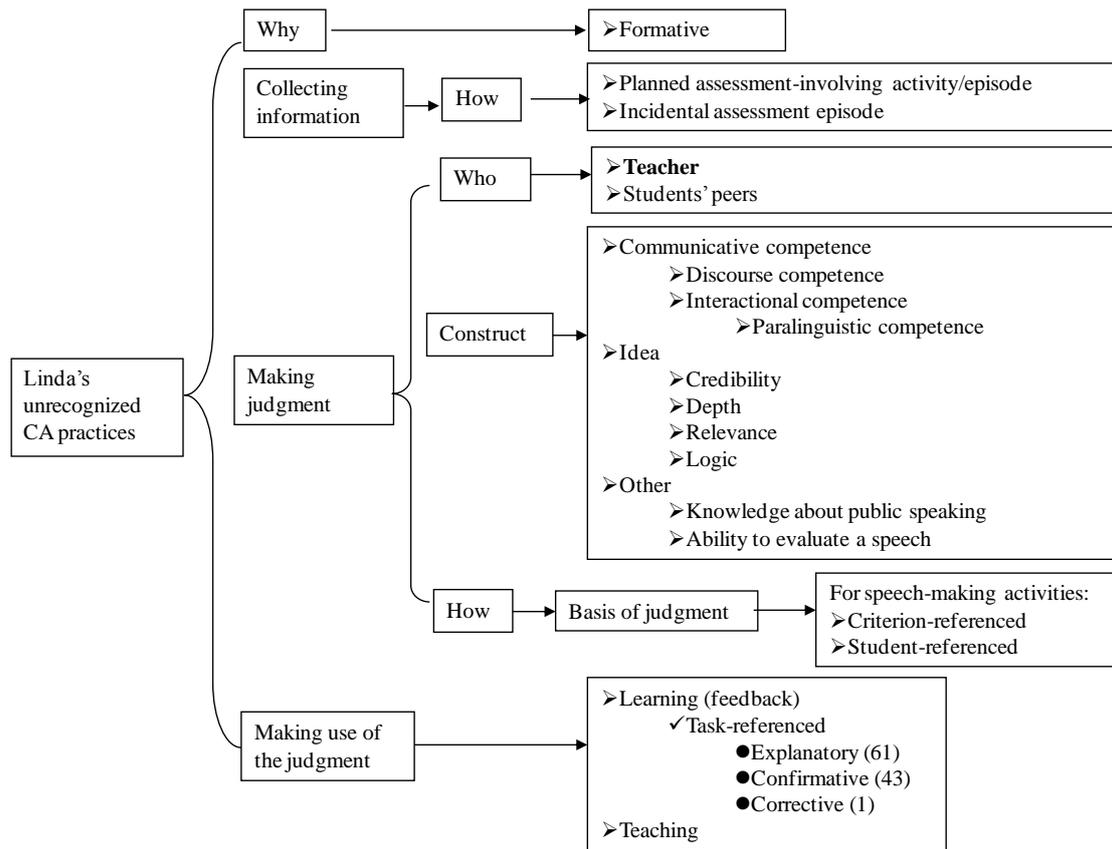
Figure 6.4 Profile of Linda's Unrecognized CA Practices

6.3 Students' general perceptions

This section will report the results of the end-of-semester questionnaire (Appendix 7) data, to triangulate with the findings from the qualitative data reported above. All the 25 students responded to the end-of-semester questionnaire but some students did not put down a response for some items. The following report will follow the structure of the questionnaire.

6.3.1 What students considered as CA

Analysis of the data from part II of the questionnaire showed that the students in this class unanimously agreed that the three prepared-speech tasks were assessment and would affect their final scores, and the majority of the students thought peer-assessment, self-assessment, instruction-embedded CA practices, and student

attendance were all assessment, but their opinions varied concerning whether they affected their final grades (see Table 6.12).

Table 6.12 What Linda's Students Considered as CA

When you were involved in the following activity:	Was your teacher assessing you? (N=25)		Would your performance affect your final score? (N=25)	
	Yes	No	Yes	No
Arriving at the classroom on time	17	8	16	9
Giving a news report at the beginning of a class	22	3	16	9
Answering your teacher's questions raised during her lecture	18	5	12	11
Presenting your ideas during a group activity while your teacher was around listening	22	3	10	15
Presenting your ideas before the whole class after a group activity	13	12	13	12
After giving a speech, responding to the questions from your classmates and your teacher	16	9	16	9
Giving three prepared speeches before the whole classes as required	25	0	25	0
Giving comments about your classmates' performances	17	8	11	14
Reflecting on your own strengths and weaknesses in giving a speech	19	6	15	10

This finding is largely consistent with Linda's understanding and practice of assessment, because she, as discussed before, only considered those CA practices that contributed to the final grading as assessment and for those instruction-embedded CA practices and student peer-assessment and self-assessment, she used them mainly as a tool to help with student learning and foster good learning habits or strategies.

6.3.2 Students' perceptions of achievement

Questionnaire data revealed that the majority of the students felt they had made some progress (see Table 6.13). Specifically, most of them felt they had improved their general oral English ability (items 1 to 6) and public speaking ability (items 9 & 10), enriched their knowledge about public speaking (items 7 & 8), and enhanced their self-efficacy (items 11 to 14) and confidence (item 15 & 16). While their performance approach goals were perceived to have been enhanced (items 17 & 18), their performance avoidance goals were perceived to have been weakened (item 19). Moreover, they felt they put in more effort to improve their oral English in general and

their public speaking skills in particular (items 20-22). The following table presents the number of students who made a specific choice.

Table 6.13 Linda's Students' Perceived Achievement

Compared with the beginning of the semester, in terms of	yours has become / is ____. (N=25)				
	much better / stronger	a little better / stronger	just the same	a little worse / weaker	much worse / weaker
1 pronunciation and intonation when speaking English	2	17	6	0	0
2 fluency when speaking English	6	16	1	1	0
3 vocabulary size of my oral English	2	20	3	0	0
4 accuracy of my oral English	1	21	1	1	0
5 complexity of the ideas that I can express in English orally	4	16	3	1	0
6 communicative ability when I talk with others in English	4	14	6	1	0
7 understanding the skills and techniques of giving a public speech in English	19	6	0	0	0
8 understanding how to improve my public speaking ability step by step	19	6	0	0	0
9 my ability to give a public speech	18	7	0	0	0
10 my ability to debate in English	15	10	0	0	0
11 the belief that I will be able to speak excellent English in future	5	16	3	0	1
12 the belief that I will become a very good public speaker in future	3	15	7	0	0
13 the belief that I will make progress so long as I work hard	10	14	0	1	0
14 the belief that I will not improve no matter how hard I try	0	1	3	12	8
15 the confidence in giving a public speech in English	10	13	2	0	0
16 the anxiety caused by giving a public speech in English	0	1	8	9	7
17 the hope to get a good score in this course	11	11	3	0	0
18 the hope to be a good student in the teacher's eyes	8	7	7	3	0
19 the worry to get a poor score in this course	0	5	10	7	3
20 the effort to create opportunities to practice my public speaking skills	4	11	10	0	0
21 the effort to enlarge my oral English vocabulary	6	12	6	1	0
22 the effort to improve the accuracy of my oral English	6	17	2	0	0

6.3.3 Students' perceptions of the causes of their achievement

Analysis of the data from part IV of the questionnaire designed particularly for Linda's class (Appendix 7) showed that three students made exactly the same choices for every item in this section, and therefore they were excluded from the analysis, making the total number for this part 22. The following table presents the frequencies of students who reported their perceptions of Linda's classroom practices, how each practice helped with their learning and to what degree. The numbers in brackets are the number of students who had made a particular choice. The four letters "a, b, c, d" in the "usefulness" column represent the four aspects listed in the "aspect" column. For the "practice" column, only those numbers that are less than the total number 22 are indicated.

Table 6.14 Linda's Students' Perceptions of the Causes of their Achievement

| Practice (N=22) | Aspect | | | | | Usefulness | | |
	Increase knowledge (a)	Develop skill (b)	Boost confidence (c)	Enhance motivation (d)	Other	High	Medium	Low
1 Explain the learning objectives of the whole course.	(7)	(11)	(10)	(16)		a: (5) b: (6) c: (5) d: (8)	a: (2) b: (5) c: (5) d: (7)	d: (1)
2 Explain the learning objectives of a particular unit. (21)	(12)	(14)	(6)	(13)		a: (8) b: (11) c: (5) d: (9)	a: (3) b: (3) c: (1) d: (3)	a: (1) d: (1)
3 Explain the learning objectives of a particular activity. (21)	(6)	(10)	(8)	(16)		a: (6) b: (7) c: (7) d: (11)	b: (3) c: (1) d: (4)	d: (1)
4 Explain the connections between the learning objectives of the whole course, of a particular unit, and of a particular activity. (17)	(6)	(10)	(10)	(12)		a: (4) b: (6) c: (6) d: (7)	a: (2) b: (4) c: (4) d: (4)	d: (1)
5 Explain the skills to be grasped in this course.	(12)	(20)	(11)	(10)		a: (9) b: (15) c: (8) d: (6)	a: (3) b: (4) c: (3) d: (4)	b: (1)
6 Ask a student to give a news	(17)	(16)	(14)	(9)		a: (11)	a: (6)	

231

Item								
report at the beginning of a class. (21)						b: (9) c: (8) d: (6)	b: (7) c: (6) d: (3)	
7 Ask questions to check how well students have learned. (21)	(16)	(10)	(8)	(12)		a: (3) b: (3) c: (1) d: (3)	a: (11) b: (6) c: (6) d: (8)	a: (2) b: (1) c: (1) d: (1)
8 Organize group activities for students to practice public speaking skills.	(11)	(22)	(18)	(12)		a: (10) b: (17) c: (15) d: (8)	a: (1) b: (4) c: (3) d: (4)	b: (1)
9 Give guidance while students practice in groups.	(13)	(20)	(11)	(10)		a: (8) b: (13) c: (6) d: (7)	a: (4) b: (6) c: (4) d: (2)	a: (1) b: (1) c: (1) d: (1)
10 Ask some students to give a practice speech in front of the class.	(10)	(16)	(14)	(14)		a: (7) b: (13) c: (10) d: (8)	a: (3) b: (3) c: (3) d: (6)	c: (1)
11 Ask students to comment on a speech given by their classmate.	(11)	(20)	(14)	(12)		a: (10) b: (16) c: (9) d: (10)	a: (1) b: (4) c: (4) d: (2)	c: (1)
12 Give feedback on one student's speech.	(16)	(18)	(14)	(10)		a: (12) b: (16) c: (9) d: (8)	a: (4) b: (2) c: (5) d: (2)	
13 Explain the marking criteria for peer assessment.	(15)	(19)	(10)	(14)		a: (12) b: (12) c: (7) d: (8)	a: (3) b: (7) c: (3) d: (6)	
14 Ask every student to give three required speeches during the semester. (21)	(17)	(21)	(17)	(14)		a: (15) b: (17) c: (13) d: (10)	a: (2) b: (4) c: (4) d: (4)	
15 Ask part of the class to give a mark for one student's speech. (21)	(10)	(17)	(10)	(15)		a: (7) b: (10) c: (6) d: (9)	a: (2) b: (6) c: (3) d: (4)	a: (1) b: (1) c: (1) d: (2)
16 Ask students to give written feedback to one classmate's speech.(18)	(10)	(18)	(9)	(9)		a: (7) b: (12) c: (5) d: (6)	a: (2) b: (5) c: (3) d: (2)	a: (1) b: (1) c: (1) d: (1)
17 Ask the class to reflect on his/her own performances after giving a speech. (20)	(9)	(17)	(14)	(13)		a: (8) b: (10) c: (10) d: (9)	a: (1) b: (6) c: (4) d: (4)	b: (1)
18 Explain to students how they will be assessed in this course. (21)	(6)	(11)	(8)	(18)	1	a: (4) b: (7) c: (4) d: (11) e: (1)	a: (2) b: (4) c: (4) d: (7)	
19 Inform students what they are expected to do at the course test(s). (21)	(10)	(15)	(11)	(10)		a: (5) b: (6) c: (7)	a: (5) b: (9) c: (4)	

					d: (2)	d: (8)	
20 Ask students to record their own prepared speeches (1)		(1)	(1)	(1)	b: (1) c: (1) d: (1)		
21 Set time limit for each prepared speech (1)		(1)			b: (1)		

It can be seen from the above table that, regarding student perceptions of Linda's classroom practices, almost all the students had responded to every item in the questionnaire, except that four students did not answer item 16, and five students did not answer item 4. Two students even added two practices (items 20 & 21). This indicated that students in general recognized and acknowledged Linda's CA practices, including explaining learning objectives (items 1-5, 18, 19), conducting classroom-embedded CA practices (items 6-10, 12, 14), and organizing peer-assessment and self-assessment (items 11, 13, 15-17). It should be noted that "student news report" activity (item 6) was not considered as a CA practice, but since this was a salient activity in this course, it was included in the questionnaire.

Regarding how each practice helped with student learning, the results showed that every practice was helpful to some students in some way, including the "news report activity". To reveal the general tendencies, the author highlighted those frequencies in the "aspect" column that were above two thirds of the total responses, that is, equal or above 16. Then it can be seen that most students felt that four practices were particularly helpful in increasing their knowledge (items 6, 7, 12, 14). While the "news report" activity (item 6) could help enrich students' background knowledge of the world, the other three practices, namely, teacher questioning (item 7), teacher feedback on a student's speech (item 12), and the prepared-speech tasks (item 14) could all help student increase their knowledge about public speaking. Most students felt that 12 out of the 19 practices were particularly helpful in improving their public speaking skills (items 5, 6, 8-17). While most of these practices were classroom-embedded CA practices (items 8-13, 15-17) and the prepared-speech tasks (item 14), the students also felt that Linda's explanation of the skills to be grasped in this course (item 5) and the news report task (item 6) helpful in improving their public speaking skills. Two practices were found particularly helpful in boosting student confidence: student group activities to practice giving a speech (item 8) and the three prepared-speech tasks (item 14). Three practices were found effective in enhancing student motivation, namely,

explaining the learning objectives of the whole course (item 1) and of a particular activity (item 3), and explaining how students were to be assessed (item 18). Overall, most students found the three prepared-speech tasks most beneficial, which corroborates with the finding from the qualitative data analysis (cf. Section 6.2.2).

Concerning to what degree such practices were helpful, it can be seen from table 6.14 that most students felt that these practices were useful at least to a medium level or even to a high level. Concerning the most frequently mentioned practices, those highlighted in the "aspect" column in Table 6.14, most students found them highly useful, except for the practice of teacher questioning (item 7), for which, most students found its usefulness just to a medium level (see the highlighted parts in the "usefulness" column in Table 6.14).

Ten students wrote down additional comments on this course. In general, their comments were positive, though three of them also expressed their wishes. They all said they had "achieved a lot" in this course, such as the improvement of "public speaking skills," "logical analysis ability," and "the ability to do research on a topic," the "strengthened confidence of speaking publicly," and the "broadening of knowledge." One student particularly mentioned the "three prepared-speech tasks", saying that "I get a very clear sense of improvement after doing the three prepared speeches, because for one thing they helped me become more confident and for another they helped me develop a clear idea about how to organize the content of a speech and how to deliver a speech". This corroborated with the finding that the three prepared-speech tasks were the most beneficial activities for the students.

The wishes from the three students were similar. They all wished to "have more opportunities to practice impromptu speeches in class" so as to "sharpen their mind" and "improve our ability to improvise opinions on an issue in precise and concise language."

6.4. Summary

In this course, Linda was engaged in a variety of assessment practices. In addition to such formal assessment as the final test and the prepared-speech tasks, she also

conducted a large number of CA practices during her everyday teaching, though she did not use such activities to grade students. Her unrecognized CA practices were embedded within three types of her classroom activities: Q&A sessions, speech-appreciation activities and speech-making activities. This group of CA practices were mostly planned assessment-involving instructional activities and episodes, but there were also some incidental assessment episodes.

Linda's CA practices can be classified into three groups: the final test which served a summative function only, the three prepared-speech tasks which served formative, summative and washback functions, and the classroom-embedded assessment practices which served formative functions only.

It turned out that students found the three prepared speeches most beneficial for their learning and improvement in comparison with the classroom-embedded FA practices or the final test. This might be attributed to the following reasons. On the one hand, these assignments were marked and students took them seriously. At the same time, Linda turned these tasks into learning opportunities by clarifying the learning intentions and criteria for success, allowing students enough time for peer feedback, requiring them to do self-reflection, and providing feedback that moved them forward. In addition, the fact that there were three such tasks instead of just one also allowed students to apply what they had learned from the previous task to the following one. In this way, students could improve through cycles of practicing and feedback and internalizing. Therefore, it might be proper to say that to some extent, the three assignments, together with the way they were conducted, achieved some kind of synergy between SA and FA. This was proven by the fact that most students found the other FA practices less beneficial because they were just a kind of "exercise" and sometimes the students did not pay enough attention (e.g., Henry_EoSI).

Regarding the specific features of Linda's CA practices, the analysis showed that Linda emphasized student involvement and student improvement in her assessment. Clearly, Linda emphasized peer evaluation and students' self-reflection in the three prepared speeches, but peer evaluation was also conducted during her speech-making classroom activities (cf. Section 6.2.4.2). This was because she wanted to "create a

learning environment" and "push the students to internalize the criteria of a good speech" (Linda_EoSI). In addition, Linda encouraged feedback that focused on the progress a student had made or how to make further improvement, which greatly enhanced some students' sense of achievement (e.g., Lewis_EoSI).

Regarding the assessment constructs, Linda's CA practices assessed students' knowledge about making a speech, ability in evaluating a speech, and ability in making a speech, with great emphasis put on the last one. Through the prepared-speech assignments, the final test, and the classroom speech-making activities, she emphasized the idea of a speech and the delivery skills, but did not pay much attention to students' language use, especially during her FA practices. By involving the students in peer evaluation and self-critique, she also wanted to promote students' ability in evaluating a speech, but she did not evaluate such abilities.

Moreover, unlike that in Andrew's class or in Mary's class, Linda's final test was not confined to her own class. Instead, students from the whole grade took the test at a fixed time and followed the same procedures. Many other teachers were also involved in the assessment. These could be considered as moderation procedures to ensure its quality.

Finally, a high degree of alignment was evidenced in this course between course objectives, course content, and Linda's CA practices. What was running through this alignment was Linda's belief that "helping students learn was more important than grading students" (Linda_EoSI).

Chapter 7. Comparative Analysis and Discussion

This thesis has investigated the complexity and dynamics of three tertiary-level EFL teachers' CA practices in their oral English classrooms in China. The researcher has not only described the recognized but also the unrecognized CA practices in each context, both cross-sectionally and longitudinally. While the research findings from the individual cases have been presented in the three previous chapters, this chapter will focus on the similarities and differences among the three cases. As can be seen from the results chapters, teachers' explanations and students' perceptions were integrated with the discussion of relevant CA practices. Accordingly, in this chapter, the comparative analysis and the discussion are also centred on the teachers' actual CA practices, with teachers' explanations and students' perceptions integrated with specific CA practices. Specifically, based on the descriptive framework (Figure 3.3), the common features concerning specific aspects of the three teachers' CA practices are presented and discussed first, followed by the discussion of the dimensions along which the three teachers' CA practices varied. The dynamics of the identified CA practices, as reflected through the relationship between SA and FA practices, are presented in the third section of this chapter.

7.1 Common features of the three teachers' CA practices

The three teachers were teaching in different educational contexts: from a comprehensive university to a foreign language university, from teaching English majors to non-English majors, from a "high structure" environment to a "low structure" environment (Wette & Barkhuizen, 2009, p. 198), and from working as an experienced teacher to working as a new teacher in a new teaching context. In spite of such differences, the three teachers demonstrated many common features regarding their CA practices.

7.1.1 Common features concerning assessment purposes / uses

7.1.1.1 Three categories of CA practices

In this study, the three teachers' CA practices fell into three categories. They not only conducted CA practices that served summative or formative purposes solely but also had CA practices that were used to help with student learning and/or improve teaching and at the same time contributed to the final grading (Andrew's presentation task; Mary's impersonated-speech task; and Linda's prepared-speech tasks). This third category corroborated the research finding that the same assessment procedures can serve both formative and summative purposes (Carless, 2011; Rea-Dickins, 2001, 2006; Rea-Dickins & Gardner, 2000).

This study also found a match between the three categories and their respective functions, namely, SA serving reporting purposes only, FA serving only learning and teaching purposes, and those dual-purpose CA serving both reporting and learning, and sometimes teaching purposes as well. This finding was understandable since the two terms are defined from the perspective of assessment purposes, which incorporate both the intention and the actual effects (Black & Wiliam, 2009). Besides, regarding specific types of feedback for learning purposes, the study found that the three teachers rarely gave person-referenced feedback. Their preference for task-referenced feedback showed that they focused on students' actual performances rather than on how intelligent they were.

7.1.1.2 Obliviousness but prevalence of FA practices

As mentioned in the previous three chapters, each teacher's understanding of CA was restricted to those practices that contributed to the final grading, those salient forms of CA practices, and they did not regard those assessment-involving classroom activities and classroom interactions as assessment (cf. Sections 4.2.1, 5.2.1, 6.2.1). However, discourse analysis of classroom recording data together with classroom observation field notes revealed that over half of the three teachers' class time was spent on assessment-involving activities (cf. Sections 4.2.4, 5.2.4, 6.2.4). This corroborated with Stiggins and Conklin's (1992) finding that a typical teacher can spend as much as

one third to a half of their class time being involved in assessment-related activities. Such a finding also supported Black and Wiliam's (1998a) finding that teachers were constantly engaged in CA during their classroom teaching, though sometimes they might not be fully aware of that.

7.1.1.3 Dual-purpose CA practices most beneficial

This study found that, comparatively speaking, the dual-purpose CA practices (Andrew's presentation task; Mary's impersonated-speech task; and Linda's prepared-speech tasks) were more beneficial for students than the other two categories. This finding could be accounted for from the following two aspects.

On the one hand, students took the dual-purpose CA practices more seriously than instruction-embedded FA practices because of the grading element incorporated in the dual-purpose CA practices. Consequently, students put more effort into such activities, as shown in the data. In Andrew's case, classroom observation showed that every presenter at the beginning of each lesson was well prepared. Most of them had prepared power point slides and speaking notes, and some of them even wrote out the whole script for their presentation. However, when Andrew organized some classroom activities, such as a role-play activity, some students would chat in Chinese. In Mary's case, regarding the midterm test, Mary's chat with some of the presenters at the beginning of that day's lesson showed that the three students who gave their impersonated speeches had practiced their speeches more than five times and had looked up all the new words in the speeches. One of them even looked up for some background information about her impersonated speech. For those students who gave their own speeches, classroom observation showed that they had all prepared some speaking notes or scripts as well as some objects, and one of them also prepared a power point presentation. However, when it came to everyday classroom activities, for example, at the beginning of the story-telling activity in the week 8 lesson, when Mary asked the students to tell their stories in pairs first, several students began to use their mobile phones to search for a scary story. Obviously, they did not prepare for this task. In Linda's case, although she conducted many speech-making activities during her daily classroom teaching, the seven volunteer students in general felt that such

classroom activities were not as helpful as the three prepared speeches because such activities were "not graded", "not that formal", and the students "did not have to prepare for them" (Helen_EoSI, William_EoSI, Henry_EoSI).

On the other hand, unlike the final tests from which students got little feedback, students could get immediate and detailed feedback from these dual-purpose CA practices, which could guide their further study (cf. Sections 4.2.2, 5.2.2, 6.2.2). Consequently, students felt they had learned most from such dual-purpose CA practices, especially so in Linda's class. Such dual-purpose CA practices in general not only helped students improve their oral English skills but also enhanced their motivation and self-efficacy.

Such a finding indicates that in this EFL context dual-purpose CA practices are more beneficial than those that serve only one purpose. While Carless (2011) has reported the effort to make formative use of summative tests in Hong Kong, the present study showed that such practices are already happening in this Chinese EFL context and do have positive effects on student learning. This finding demonstrates the combined positive effect of the inherent washback effect of SA on students and the usefulness of FA for student learning.

7.1.2 Common features concerning assessment methods

7.1.2.1 Three types of eliciting methods

This study found that every teacher employed a variety of assessment methods that fell into the following three categories: *formal assessment*, usually in the form of midterm/final tests and marked assignments; *planned assessment-involving instructional activities/ episodes,* usually in the form of classroom activities or interactions that were planned in advance and involved students' performances in some way, such as doing a role-play or asking a prepared question; and *incidental assessment episodes,* often in the form of scaffolding questions or comments during classroom interactions. This study also found that sometimes some incidental assessment episodes were embedded within planned assessment-involving instructional activities or episodes.

Such findings to a great extent supported Cowie and Bell's (1999) model of FA (cf., Figure 2.1) in that 1) *planned assessment-involving instructional activities* were similar to planned FA in their model, 2) *incidental assessment episodes* were similar to interactive FA in their model, and 3) the fact that sometimes *incidental assessment episodes* were embedded within *planned assessment-involving instructional activities* was similar to the connection between planned and interactive FA through purpose. However, the *planned assessment-involving instructional episodes* were not accounted for in their model.

7.1.2.2 Patterns of planned assessment-involving instructional activities / episodes

Regarding the CA practices embedded within classroom teaching and learning, this study found one typical pattern for the planned assessment-involving instructional episodes, and one typical pattern for the planned assessment-involving instructional activities.

The planned assessment-involving instructional episodes often took the form of the IRF (teacher initiation – student response – teacher feedback) sequence (Sinclair & Coulthard, 1975) or its variations, such as the omission of the initiation step (e.g., Excerpt 4.1). This pattern usually occurred during such input-oriented activities as teacher lectures (in Mary's case), listening comprehension activities (in Andrew's case), and reading-comprehension activities (in Linda's case). This type of IRF sequences were usually short, sometimes embedded in the teacher's teaching discourse (in Mary's case) and sometimes one IRF sequence after another in an activity depending on the number of questions asked (in Andrew's and Linda's cases). This type of IRF sequence reflects the typical features of "convergent assessment" (Torrance & Pryor, 2001, p. 617) because it aimed to discover if students knew or understood something predetermined by the teacher. The fact that all the teachers used IRF sequences reflected the fact that in these oral English courses, the teachers were concerned about whether students were following them during some language- or knowledge-input activities. This is understandable since oral English courses are not just about asking students to talk in English. Students need input before they can

produce output, and the short IRF sequences could help teachers identify students' problems in taking in the input and provide help immediately.

The teachers' planned assessment-involving instructional activities often took the form of "teacher initiation – student practice – student demonstration – teacher / student feedback," or IPDF in short. Here *student practice* and *student demonstration* are two distinct steps. *Student practice* refers to the step when students do the assigned tasks individually, in pairs, or in small groups, while *student demonstration* refers to the step when the teachers asked individual students to perform before the whole class. In the *student practice* step, every student should be doing the assigned task, while in the *student demonstration* step, only a few students would have the chance to do the task again while the rest of the class would watch and listen.

This IPDF pattern is different from the IRF sequence in three ways. First, while the IRF sequences usually occurred during input-oriented activities, this IPDF pattern and its variations (e.g., Figure 5.4), usually occurred during the output-oriented activities such as Andrew's role-play activities, Mary's story-telling activities, and Linda's speech-making activities. Second, while the IRF sequences were usually very short, often less than one minute each, an activity involving the IPDF pattern usually took longer than 10 minutes and sometimes as long as 80 minutes, as in the case of Mary's story-telling activity. Moreover, while the IRF sequences were mainly concerned with whether students knew or understood a predetermined thing, the IPDF showed the typical features of "divergent assessment", aiming to "discover *what* the learner knows, understands, and can do" (italicized original) (Torrance & Pryor, 2001, p. 617).

The fact that all the teachers used the IPDF patterns or its variations, which occupied a large amount of class time, indicated that all the teachers were primarily concerned with providing students with opportunities to practice their oral English and offering immediate feedback to help them improve. That there was a distinctive step of student demonstration was probably due to the big class size and the limited class time, which made it impossible for every student to perform before the class. However, this step was indispensible because the teachers needed some students' performances to serve

as models so that they could point out some common problems, which could also benefit other students (Andrew_EoSI, Linda_EoSI).

7.1.3 Common features concerning the judgment-making process

7.1.3.1 Teacher as the major agent of judgment

This study found that in each context, the teacher was usually the most important agent of judgment. This feature was obvious in both Andrew's case and Mary's case, where most feedback in both recognized and unrecognized assessment practices was provided by the two teachers rather than their students or any external agent, although the teachers occasionally did invite the class to make some comments about other students' performances. In Linda's case, while she conducted student peer assessment and self-assessment more frequently, such practices were always under her guidance, and she was always a most significant person in making judgment and providing feedback, in both the marked assignments and in those assessment-involving classroom activities. Therefore, the teachers played the most significant role in the judgment-making processes of CA.

7.1.3.2 Assessment constructs: More than oral English abilities

While the three courses were all oral English courses and the constructs of the three teachers' CA practices were mainly about students' communicative competence, the study found that the teachers assessed more than just students' oral English abilities. Their additions can be divided into two groups.

First, they all included aspects that were academically related to the course content or the course learning goals but not included in the comprehensive model of communicative competence (Figure 2.12). Specifically, they all assessed the quality of students' ideas and their use of visual /aural aids. In addition, Andrew also assessed students' cooperation skills, Mary also assessed students' knowledge about western culture, and Linda also assessed students' knowledge about public speaking and their ability to evaluate a speech. While these aspects were not part of students' oral English

abilities, they were closely related to each teacher's daily teaching and/or course learning goals.

Second, they all assessed some non-academic factors. For example, they all assessed *student attendance* and *class participation/ interaction*, although what constituted *class participation / interaction* was open to interpretation, as can be seen from Mary's students' various interpretations (cf. Section 5.2.4.4). These two aspects were clearly included in Andrew's and Linda's respective formal assessment structures (cf. Appendixes 12 & 21). In Mary's case, though she did not specify the actual percentages of these aspects, she indicated in her interviews that if some students were late or absent too often, she would make a note of that and their final scores would be affected (Mary_SRI). In addition, Mary also looked at students' attitude towards learning (cf. Table 5.10); Linda also looked at students' effort (cf. Section 6.2.4.2).

This feature suggests that the teachers were not only concerned about course content and course learning goals, but also about many other aspects, including non-academic factors. Their final grading was not purely about student learning. This to some extent reflected the "hodgepodge" nature of teachers' grading practices (Brookhart & Nitko, 2008).

7.2 Variations concerning the three teachers' CA practices

While the three teachers showed some common features regarding assessment purposes and assessment methods, when it came to the specific steps of the assessment process, variations emerged among the three teachers, especially relating to their eliciting methods in FA practices, the agent of judgment, the basis of judgment, and the major constructs of their assessment.

7.2.1 Frequency of incidental and planned FA practices

This study found that the three teachers varied greatly regarding the frequency of their *incidental assessment episodes*. While such practices often occurred in Mary's case (altogether 181 instances in the 313-minute observed lessons), they were rare in Andrew's case (just 5 instances in the 400-minute observed lessons) and only

sometimes occurred in Linda's case (22 instances in the 358-minute observed lessons).

Such variations were found to be related to each teacher's beliefs about the best way to help improve students' oral English as well as their respective course content. As Mary believed that the best way to learn and improve oral English, especially speaking fluency, was through engaging in authentic communication (Mary_EoSI), she often assigned students tasks which had authentic communication needs such as sharing personal experiences or opinions and she often acted as a true interlocutor in class by having extended conversations with individual students. Moreover, her course did not follow any textbook and consequently she did not have the obligation to finish certain tasks by a certain time. Naturally, her classroom interactions with the students were more like natural conversations and her feedback was more spontaneous (cf. all the excerpts in chapter 5).

In contrast, neither of the other two teachers adopted such a view of oral English learning. Instead, Linda emphasized 1) one's thinking as reflected through one's speaking, which was also specified in the course description (cf. Appendix 21), and also 2) that students should take responsibility for their own learning. Thus, during the feedback sessions, she often engaged students in analyzing and evaluating a sample speech or students' own speeches to help them internalize the characteristics of good speeches as well as to practice their thinking (e.g., Excerpt 6.3). Therefore, sometimes incidental FA practices occurred. However, Linda's course was specifically devoted to public speaking skills and was guided by two textbooks, which unavoidably provided a structure for the course and conditioned her teaching to be pre-planned. Consequently, many of her classroom activities were also planned assessment opportunities.

As for Andrew, he believed that to learn oral English, students have to "first imitate the way native speakers talk" and then "practice a lot" (Andrew_EoSI). Therefore, the teacher's role in class was to provide "the models" for students to imitate and offer many opportunities for students to practice (Andrew_EoSI). Since Andrew took his classroom activities mainly as practice opportunities, it was understandable that he

was seldom engaged in scaffolding to help students give a better performance, since he was just giving students a chance to practice. Besides, Andrew as a new teacher followed the textbook given to him strictly. Since the textbook already contained many activities and Andrew needed to finish one unit in two sessions, his major task turned out to be to finish the tasks listed in the textbook. Therefore, the great majority of his FA practices were planned, because they had been planned in the textbook.

Since incidental FA practices should enable the teachers to "adjust their teaching in light of students' responses to questions or other prompts in 'real time'" (Wiliam & Thompson, 2008, p. 70), the differences mentioned above among the three teachers revealed that Mary's FA practices were highly responsive to students' needs, and Andrew's were least responsive. However, Mary's students did not feel they had learned much in class. By contrast, Linda's students felt they had learned a lot. This was probably because Mary's course lacked a clear learning goal while Linda's course had clear and specific learning objectives, and throughout the semester, Linda kept reminding students of the learning goals in one way or another. This showed that "clarifying and sharing learning intentions and criteria for success" is an important strategy for FA (Wiliam & Thompson, 2008, p.64).

7.2.2 Transparency of CA to students

The three teachers also varied in terms of how transparent their CA practices were to their students. Linda's CA practices were found the most transparent compared to the other two teachers'. For one thing, she notified her students of when, what, and how they would be assessed in this course both at the very beginning of the semester, through a detailed course description, and a couple of weeks before each scheduled assessment task. For another, she made clear to students the criteria for success, through the specific marking criteria, and organized self-assessment and peer-assessment to help them internalize such criteria. Her formal assessment structure, except for the "class interaction" part, which remained vague to students, was generally very transparent to her students.

In Andrew's case, while the final test and the presentation task was made clear to students at the beginning of the semester through the course description, the "class

participation" component, which accounted for 30% in the final grading according to the course description, remained vague to students as well as to Andrew. Even at the end of the semester, different students still had a different understanding of what this part actually meant. Andrew, a new teacher in his teaching context, had to work out his own way to interpret this ambiguity (cf. Section 4.2.1) since he was just given the course description at the beginning of the course and had no input into what it stated about the assessment components. Therefore, his CA practices were less transparent than Linda's.

In contrast, Mary's CA practices were the least transparent to students compared to the other two teachers'. Throughout the semester, students were not very clear about the course objectives. They had a vague idea that this course was to help them improve their oral English, but they were unsure about which aspects and how. Besides, students were informed of the midterm test and the final test only about a couple of weeks beforehand. They were told about when, how, and which topic they would be assessed on, but were not given any criteria for success or informed of how their performances would be evaluated. Therefore, Mary's CA practices were the least transparent to students.

A comparison of such practices against Wiliam and Thompson's (2008) model of FA/AfL (cf. Figure 2.6) and Davison's model of TBA (cf. Figure 2.7) revealed that Linda's practices had more of those ideal features presented in the two models than the other two teachers'. Most importantly, she made the learning goals and criteria for success clear to students and constantly involved students in activities through which they could understand the goals and criteria better. It was in these two aspects that the other two teachers lagged behind.

Since the three teachers all showed great enthusiasm towards teaching and none of them were novice teachers in teaching oral English, such differences could to some extent be attributed to their beliefs and knowledge about assessment and their past assessment experiences.

In Andrew's case, when asked if he would stress the marking criteria of an assessment task so that students would use them as their learning goals and strive hard to achieve them, he said,

> I won't. I don't want to use tests to make my students work harder. Actually, I have always been telling them that you should make the learning process an enjoyable process. Don't focus on the tests. Otherwise, you will become too goal-oriented and you will miss the fun of the learning process. (Andrew_EoSI)

Clearly, Andrew regarded assessment as testing and thought that emphasizing the marking criteria was a de-motivating practice for students. Moreover, as a teacher new to this teaching context, Andrew had no experience of how students should best be assessed in this course and could only do what he was asked to do (Andrew_EoSI). He was not even sure about the learning goals of this course (Andrew_EoSI). Obviously, his limited view of assessment as well as lack of experience in his teaching context contributed to the fact that his CA practices were opaque to his students.

Similarly, Mary regarded assessment as grading and thought it "a necessary evil" (Mary_SRI, Mary_EoSI). Meanwhile, she thought oral English should be spontaneous talk and she associated oral English assessment with formal proficiency tests; therefore, she saw no necessity or value in informing students of what and how they would be assessed, let alone providing the marking criteria. Thus, this limited view prevented her from seeing assessment as a useful tool in class to help with teaching and learning and consequently made her CA practices opaque to her students.

In contrast, Linda believed that students should be clear about what they were going to learn at the beginning of the semester, and throughout the semester she could use assessment, or rather grading, to motivate students to work hard.

> It [assessment] can be a motive for them to learn. Particularly in China, the students are kind of conditioned to think that scores are very important and they would work for the scores. So if you say you are going to get a score for this, probably they are going to take it more seriously than something when you say no it's not going to be scored. (Linda_EoSI)

Moreover, she had a strong belief in the value of peer-evaluation:

> I think peer evaluation is important and valuable in students' learning. ...When one is asked to give comments to other people's performance, he has to listen very carefully, set criteria in his mind, evaluate the speech according to these criteria, and then give comments. This process can facilitate their learning. ... Asking them to comment on each other's work could also give them a feeling that the class is a learning community in which they help each other. (Linda_SRI)

With such beliefs, as well as her past teaching experience of this course, it was understandable that she could make her CA practices more transparent to her students.

In addition, such differences might also be related to the teachers' specific educational background and assessment experiences. Linda, with the highest degree among the three, was the most informed about assessment, and Andrew, with a Masters degree in American literature, was the least informed. Besides, Linda had the longest teaching experience, and Andrew the shortest. Consequently, Linda could be expected to be more experienced in doing classroom assessment than Andrew. This showed that more knowledge about assessment and richer teaching/ assessing experience should enable a teacher to use assessment more effectively to help with teaching and student learning.

The above analysis showed that teacher beliefs about assessment, together with their educational background and past teaching / assessment experiences, could affect their CA practices. This finding to some extent corroborated previous finding that teachers' past teaching experiences, especially assessment experiences, play a role in their present CA practices (Cheng, Rogers, & Wang, 2008; Xu & Liu, 2009). However, few previous studies have pointed out the importance of the teacher belief factor.

7.2.3 Teachers' judgment-making orientations

All three teachers were concerned about the fairness of their grading but had taken different measures to ensure this quality. Analyzing such variability against Davison's (2004) framework about teacher-rater orientations, the study found that Andrew tended to be more oriented towards the "assessor as the interpreter of the law" position, Mary more towards "assessor as God" position, while Linda more towards "the assessor as the principled yet pragmatic professional" position.

According to Davison (2004, p.324), the typical characteristics of the "assessor as the interpreter of the law" position are that 1) the assessment process is primarily dominated by criteria, 2) the assessment is primarily based on students' performances but teachers are willing to make depersonalized accommodations, and 3) inconsistency in judgment is seen as a problem or threat to reliability. Results from the present study showed that Andrew had strong concern for consistency in marking. During his grading, he "gave separate scores for each of the specified aspects in the marking criteria" to ensure accuracy of his marking and he constantly reminded himself of "not being influenced by my general impression of a student" (Andrew_EoSI), though he did make modifications so as to make the scores correlate more with his general impressions so as to be fair to students. This reflected the struggle he went through to ensure consistency in his marking.

In sharp contrast, Mary's judgment-making process showed some typical features of the "assessor as God" position. She repeatedly mentioned that she did not have specific marking criteria, her marking was "intuitive", and actually she found analytical scoring "very troublesome and ineffective"; instead, she found holistic and impressionistic scoring efficient and also quite reliable because her scores on different occasions correlated very well (Mary_EoSI). Besides, she was primarily concerned with individual students' relative positions in this class so as to meet the university assessment policy. Such features as being "intuitive" and "community-bound" were the typical features of "assessor as God" position (Davison, 2004).

Linda was somewhere in between and more oriented towards "the assessor as the principled yet pragmatic professional" position because she showed more confidence and flexibility in her judgment-making process, as seen from her interview at the end of the semester when she talked about her difficulties in assessing students.

W: Ok. Have you come across any difficulties when you assess your students?

L: Yes, I always feel it quite difficult to give them scores.

W: Why? You want your scores to be scientific? To be objective?

L: For that part, it's ok. We can come up with some relatively objective or reasonable grading criteria. That part is Ok. But like I said, when I score them, on

the one hand, I want to be fair; on the other hand, I don't want them to get discouraged, if they get a score which is lower than they expected. So I want to give them all very high score. but if I do that, still it's going to hurt some students, because they might think it's unfair. But still for the scores, I still want to kind of encourage them with scores.

W: You want to use it as a tool to encourage students?

L: Yeah. So this is a little contradictory to the so-called objectiveness of the criteria of a score, so that's why I always feel quite difficult. (Linda_EoSI)

The underlined part showed that Linda was very confident about her marking criteria as well as her ability to mark reliably. However, she was very much concerned that marks might de-motivate students. Consequently, she inflated her scores in relation to students' actual levels in class so as to motivate students to work harder. Therefore, her judgment-making process was mainly criterion-referenced, but she also made localized accommodations to enhance student learning. These were typical features of the "assessor as the principled yet pragmatic professional" position.

The variation among the three teachers along this dimension was found related to each teacher's specific teaching context, as well as their belief about the desirable way of grading. Contextually, both Andrew and Linda were teaching comparatively bigger classes and there were parallel classes who were taking the same course during the same period of time. Therefore, the use of marking criteria could ensure a higher level of consistency across different classes during grading. As members of their respective teaching groups, they had the responsibility to follow the marking criteria. In contrast, Mary's teaching context allowed her almost total freedom as to what, when, and how to assess so long as she observed the university assessment policy that less than 40% of the students in her class got a score above 85. Also, she was the only teacher for the oral English course in her department and she did not have to keep in step with another teacher. Such freedom allowed her to choose the most efficient way to grade: to grade impressionistically and intuitively. This finding was different from what Cheng and Wang (2007) found. In their study, they found that teachers from Mainland China used holistic scoring most due to their large class sizes. In this study, the "class size" factor did not play as important a role as the "requirement from the teaching group or university" factor.

251

Meanwhile, teachers' grading practices were also found related to their beliefs about what was the best way to grade. Both Andrew and Linda believed that analytical scoring tended to ensure consistency and therefore reliability, whereas Mary thought it troublesome and ineffective. Besides, while all the three teachers hated grading students, Andrew and Mary tended to regard this part as a burden while Linda regarded it as a tool to motivate students to learn. Such different beliefs naturally led to their different practices.

7.2.4 Relationship between assessment constructs and course learning objectives concerning oral English ability

It has been pointed out that for a curriculum to be effective, there should be alignment between curriculum objectives, classroom instruction and student assessment, and this kind of alignment is usually achieved through turning the course learning objectives into assessment constructs in CA practices (English, 1992). This section will compare this kind of alignment among the three cases.

The data for this part of the analysis was restricted to each teacher's CA practices that assessed students' oral English ability only, consisting of three sets: each teacher's final test, the feedback sessions of each teacher's dual-purpose CA practices, and the feedback sessions of each teacher's FA practices concerning students' oral output (Table 7.1).

Table 7.1 Data Sources for Comparing the Relationship between Assessment Constructs and Course Learning Objectives Concerning Oral English Ability

CA type	Andrew	Mary	Linda
SA	The final test	The final test	The final test
Dual-purpose CA	The presentation task [85 minutes] [9 students were assessed]	The midterm test [107 minutes] [9 students were assessed]	The prepared-speech tasks [192 minutes] [16 students were assessed]
FA	Seven discussion activities Three role-plays [154 minutes]	Two discussion activities One warming-up chatting One story-telling activity [167 minutes]	Five speech-making activities [132 minutes]

Analysis of the three teachers' assessment constructs concerning oral English ability showed different patterns, as can be seen in Table 7.2. The numbers in the table, which are not put in brackets, are the actual frequencies of a specific assessment construct found in a specific teacher's specific type of CA practices. Those numbers that are put in brackets are the total numbers of particular categories of assessment constructs. It should be pointed out that for the dual-purpose CA practices and for the FA practices, the frequency numbers are based on the analysis of relevant classroom interaction sections. However, for the frequency numbers in the SA category, they were based on the marking criteria in Andrew's case (cf. Table 4.6) and in Linda's case (cf. Table 6.8), and were based on the limited recording data in Mary's case, because of the lack of recording data, the lack of marking criteria for Mary's case, and the lack of immediate teacher feedback to individual students after each of the final tests. In addition, in Andrew's case, the memorization element in the final test might have undermined the constructs claimed to be assessed in the test (cf. Section 4.2.3).

Table 7.2 Assessment Constructs Concerning Oral English Ability in the Three Cases

Frequency / Aspect	Andrew			Mary			Linda		
	SA	Dual	FA	SA	Dual	FA	SA	Dual	FA
Interactional competence				(1)					
Paralinguistic competence	(15)			(0)			(37)		
Body language	/	9	3	/	/	/	1	3	2
Voice volume	/	3	/	/	/	/	1	/	3
Confidence	/	/	/	/	/	/	/	10	2
Tone	/	/	/	/	/	/	/	9	/
Eye contact	/	/	/	/	/	/	1	4	/
Use of stress	/	/	/	/	/	/	/	1	/
Conversational competence	/	/	/	1	/	/	/	/	/
Linguistic competence	(17)			(84)			(11)		
Accuracy	(5)	(7)	(2)	(2)	(15)	(66)	/	/	/
Pronunciation	1	3	/	1	15	6	/	/	/
Vocabulary	2	3	/	/	/	44	/	/	/
Grammar	2	1	2	1	/	16	/	/	/
Speaking speed/ fluency	1	2	/	1	/	/	2	8	/
Use of parallel structure	/	/	/	/	/	/	/	1	/
Sociocultural competence	1	/	4	/	1	1	/	/	/
Discourse competence	(4)			(8)			(46)		
Idea progression	1	1	/	/	4	/	1	6	10
Coherence	1	/	/	/	/	/	2	13	/
Content relevance	1	/	/	/	4	/	1	9	4

	C1	C2	C3	C4	C5	C6	C7	C8	C9
Formulaic competence	/	/	3	/	/	/	/	/	/
Cooperation skills	(5)			/			/		
Team work	1	4	/	/	/	/	/	/	/
Use of aids	(5)			(1)			(3)		
Aural aids	/	3	/	/	/	/	/	/	/
Visual aids	/	2	/	/	1	/	/	3	/
Ideas	(1)			(128)			(47)		
Depth	/	/	1	1	/	26	/	10	4
Logic	/	/	/	/	/	1	1	9	3
Clarity	/	/	/	/	/	98	1	3	3
Credibility	/	/	/	/	/	/	/	2	10
Adequacy	/	/	/	/	/	/	1	/	/
Significance	/	/	/	/	3	/	/	/	/
Rhetorical effectiveness	/			/			(12)		
Tone	/	/	/	/	/	/	/	9	/
Use of humor	/	/	/	/	/	/	/	2	/
Overall effect	/	/	/	/	/	/	1	/	/

A comparison between what each teacher actually assessed during the CA practices and their respective course objectives as well as their marking criteria if available revealed that there was a very high level of alignment in Linda's case, and a fairly high level in Andrew's case, but little alignment in Mary's case, as can be seen in the following three tables. The highlighted parts are those areas where alignment was achieved.

Table 7.3 The Alignment between Linda's Course Objectives, Marking Criteria and Actual CA Practices

Course objective	Marking criteria	Linda's actual CA practices
Ability to deliver a public speech	Ideas	Ideas
	Relevance	Relevance
	Depth and logic	Depth and logic
	Clarity	Clarity
	Credibility	Credibility
	Rhetorical effectiveness	Rhetorical effectiveness
	Discourse competence	Discourse competence
	Linguistic competence	Fluency of linguistic competence
	Paralinguistic competence	Paralinguistic competence
	/	Use of aids
Ability to evaluate a speech	/	/
Enhance confidence	/	Confidence

254

It should be noted that Linda assessed students' ability to evaluate a speech by requiring them to do peer evaluation and self-evaluation during the three prepared-speech tasks. In class, she also frequently invited the students to comment on their peers' performances. Therefore, she provided many opportunities for students to practice this ability to evaluate a speech. Therefore, there was a high level of alignment in her case.

Table 7.4 The Alignment between Andrew's Course Objectives, Marking Criteria and Actual CA Practices

Course objective	Marking criteria	Andrew's actual CA practices
Enhance self-confidence	/	/
Improve fluency	Fluency	Fluency
Improve accuracy	Pronunciation, grammar, vocabulary accuracy	Pronunciation, grammar, vocabulary accuracy
Enlarge vocabulary	Vocabulary diversity	/
/	/	Formulaic competence
Increase knowledge	/	/
Enhance conversational skills Enhance presentation skills	Discourse competence Speaking with proper manners Team work	Discourse competence Paralinguistic competence Team work Use of aids
Enhance story-telling skills	/	/

In Andrew's case, he sometimes designed specific activities to enlarge students' vocabulary, but he did not particularly comment on students' vocabulary size during his CA practices.

Table 7.5 The Alignment between Mary's Course Objectives and Actual CA Practices

Course objective	Mary's actual CA practices
Improve fluency or rather student confidence	/
	Clarity and depth of ideas
	Accuracy of linguistic competence

In Mary's case, while she claimed that the course intended to help students open their mouths and speak more confidently using the language they had already learned (cf. Section 5.1.3), what she emphasized was students' language accuracy and the clarity / depth of their ideas.

Such variations in the level of alignment among the three cases may be attributed to each teacher's knowledge about English language teaching and learning, teaching and assessing experience, and their respective teaching context.

Linda's PhD degree is in the field of second language acquisition, and before the time of the study, she had taught two courses to Masters-level students majoring in applied linguistics: *Material Design and Evaluation* and *Introduction to Applied Linguistics*. During her more than 20 years of teaching, she had not only attended numerous conferences and teacher training programs but also been an oral-English examiner many times for some large-scale exams in China such as College Entrance Examination Oral English Test, and Adult Self-Study Oral English Test Band One and Band Two. As for her oral English course, she had been the coordinator for this course for over 10 years; she had witnessed its development and had been an active member in promoting changes in it. She had been teaching the course twice before she participated in this study and she had collected feedback from her previous students, both formally through questionnaires and informally through interviews, about this course and made adjustments accordingly each time. Probably her expertise in English language teaching and learning, her past teaching and assessing experiences, and the fact that her teaching context allowed her and her colleagues to make reforms, all helped her to be more and more clear about what this oral English course should be truly aimed at and how the course aims should be achieved so that the course could best serve students' needs. Naturally, there was a high level of alignment in her case.

Like Linda, Andrew was also teaching according to a group-determined course syllabus, which contained the course objectives. However, unlike Linda, he was not an active participant in the design of the course objectives and the syllabus. Instead, he was a passive receiver because he was new to this teaching context. Thanks to this group-determined course description and the required textbook, there was a fairly high level of alignment in his case because he was careful to complete the required lessons and tasks by the scheduled time, even though his knowledge and teaching / assessing experience were comparatively more limited. This indicates that a novice teacher might benefit from teaching in a "high structure" context, which might offer guidance

and help to the teacher to ensure a high level of alignment between course objectives and the teacher's CA practices.

In contrast, the low level of alignment in Mary's case might have been partly due to her "low structure" teaching context, which made it unnecessary for her to articulate her course objectives, and in turn did not force her to think carefully about what the course was aimed at. Therefore, while she claimed that the course was to enhance students' speaking fluency by pushing them to talk in class (Mary_BI), what she actually focused on in class was mainly students' language accuracy and clarity of idea (see Table 7.2). This reflected the fact that she had not realized the importance of establishing clear learning goals for a course and transmitting the goals to students through various tasks, as suggested by several FA / CA theories or frameworks (Cowie & Bell, 1999; Davison & Leung, 2009; Wiliam & Thompson, 2008).

Consequently, students in Linda's class found her course very beneficial but some of Andrew's and Mary's students found some of their teachers' practices not that useful for their learning. With the course objectives running through Linda's CA practices, her students were constantly reminded of what a good speech was like. Such CA practices were also opportunities for students to reflect on and internalize the criteria of a good speech, which guided their further efforts. Naturally, by the end of the semester, all the interviewed students felt that they had improved, both in skills and in confidence, and they were all able to articulate the criteria of a good speech as well as being clear where they should put more effort. In contrast, without clear and consistent course objectives being communicated through the teachers' CA practices, students in Mary's and Andrew's classes could only resort to their own understanding about what they should learn and how they should learn. Consequently, some of them could not fully appreciate and/or benefit from their teachers' CA practices. For example, Jack in Andrew's class felt that doing the role-plays was like giving a kind of fake talk and he did not realize that such activities could help him grasp those functional language expressions effectively (Jack_EoSI). Although Mary provided online feedback constantly in order to help students improve the quality of their oral performances, some of her students could not fully take in her feedback because they were mainly worried about the impression they might have left in her eyes, which might affect their

scores in this course. For example, while Daniel recounted his unlucky weekend during the warm-up chatting in the week 8 lesson, Mary provided many useful expressions along the way to make his talk more interesting and more correct. However, in the stimulated retrospective interview afterwards, Daniel said he did not remember Mary's corrections because in class he was busy thinking what he should say next. It can be seen that when a teacher's CA practices were not linked through the course objectives and made explicit to students, it was likely that the teacher's good intentions might not be appreciated and the teacher's efforts might not be valued.

7.3 Relationship between SA and FA

Many researchers have pointed out that while FA is obviously beneficial for teaching and learning, SA is an indispensable part of education, and synergy should be achieved between the two since in essence SA can also be regarded as FA in the long run (Brookhart, 2010; Carless, 2011; Davison, 2007; Davison & Leung, 2009; Harlen, 2005, 2006, 2007; Harlen & James, 1997). A comparative analysis of the three cases revealed that the three teachers varied according to the level of synergy between their FA and SA practices.

In Linda's case, her FA and SA practices were highly integrated and fed into each other, looking like an upward spiral centred on the course learning objectives. Thanks to the high level of consistency between her course objectives and her assessment constructs (cf., Table 7.3), her FA and SA practices were found linked to each other through the repeated emphases on the same assessment constructs. More importantly, 60% of the final grading was based on the dual-purpose CA practices and another 15% was from students' self-assessment and peer-assessment. These practices took advantage of both SA's function as a motivation tool, especially in the Chinese educational context and FA's function as a teaching and learning tool (cf., Cizek, 2010). Since her dual-purpose CA practices were integrated with her daily teaching and spread out during the second half of the semester, the feedback students got from earlier FA practices could be used in helping them to accomplish these marked assignments more successfully, and the feedback from the earlier marked assignments could be made use of by the students in their later FA practices, marked assignments,

and the final test. It can be seen that her FA and SA practices were closely linked to each other and formed a kind of spiral that ran through the whole semester, pushing students higher and higher to achieve the course objectives. Obviously, Linda's FA and SA practices achieved a high level of synergy.

In Andrew's case, his FA and SA practices were intentionally separated from each other, looking like two parallel lines, though there was the regular presentation task, which was like a tiny knot that linked SA and FA in each lesson. Thanks to the course description and the textbook, Andrew was engaged in a lot of FA practices in his everyday teaching and what he emphasized in these practices was more or less consistent with what the course was aimed at. However, probably due to a lack of teaching and assessment experience as well as a lack of knowledge about CA, Andrew did not make the two types of CA practices work for each other. Instead, he followed the traditional idea that assessment should sit at the end of a semester to evaluate student achievement in an objective and fair way, and he did not regard other classroom activities involving assessment as CA. It can be seen that the synergy between his FA and SA practices was quite low.

In Mary's case, a moderate level of synergy was achieved through the fact that part of her midterm test, the impersonated speech task, was based on her assessment of students during the first half of the semester, and her final test was an extension of the midterm test task as well as a classroom discussion activity. This kind of evolving pattern indicated that her FA fed into her SA. However, due to the lack of clear course objectives as well as her lack of awareness of CA practices, though her class was full of FA practices, they were like unchained pearls spreading all over the place, and they were not appreciated by the volunteer students who wanted more cognitively challenging tasks in class.

The above analysis shows that CA can be seen as a system in classroom contexts, where the traditional distinction between FA and SA is still valuable. Since assessment is essentially a process consisting of three steps: collecting evidence about student learning, making a judgment about such evidence, and then making use of the judgment (cf. Section 2.2), what links one CA practice with another in an educational

context is essentially the assessment constructs each CA practice centres on. The relationship between the two types of CA practices can be analyzed by looking at how each CA practice helps achieve the course goals. The level of synergy between the two types may affect the overall effect of a course.

7.4 Summary

This chapter presents the results of a comparative analysis of the three cases that have been described in the previous three chapters in detail. The analysis revealed both similarities and differences among the three cases regarding the specific aspects of CA. Finally, taking CA as a whole system, the analysis revealed that Linda's SA and FA showed a high level of synergy while Andrew's a low level of synergy and Mary's sitting in between.

Chapter 8. Conclusion

Adopting the interpretive paradigm and the case-study approach, the present study develops a descriptive framework (Figure 3.3) and portrays the complexity and dynamics of three EFL teachers' CA practices in their oral English classrooms in China. As the profiles of each teacher's CA practices (Chapters 4,5,6) as well as the similarities and differences among them (Chapter 7) have been presented in the previous chapters, this chapter will first summarize the major findings, followed by a discussion of their implications. The chapter will end with a discussion of the limitations of this study and suggestions for future research.

8.1 Major findings

Although the three teachers were teaching in different contexts in terms of course content, university type, department requirements, and class size, they showed the following common features regarding their CA practices.

While the teachers' recognized CA practices were restricted to those leading to the final grading, they also conducted many unrecognized CA practices that were embedded within their daily teaching. The recognized CA practices served summative purposes or dual purposes (both summative and formative) and the number of such practices was small. However, the teachers were also found to have spent over half of their class time engaging in unrecognized CA practices that served formative purposes mainly.

The recognized CA practices were usually in the form of formal assessment, and the unrecognized CA practices were usually in the form of planned assessment-involving instructional activity/episode and incidental assessment episode. The planned assessment-involving instructional activities often took the IPDF pattern (teacher initiation – student practicing – student demonstrating – teacher/student feedback) and was frequently used in language-output activities; the planned assessment-involving instructional episodes often took the IRF (teacher initiation – student response – teacher feedback) pattern and was frequently used during language-input activities;

and the incidental assessment episodes often took the form of scaffolding questions or comments and were often embedded within teacher-student interactions.

The three teachers were all the major agents of judgment in both recognized and unrecognized CA practices, although Linda conducted more of student peer-assessment and self-assessment comparatively speaking. The teachers all focused on certain aspects of students' communicative competence, but at the same time, they also assessed other elements that were related to their specific courses but not part of communicative competence, as well as the quality of students' ideas and their learning attitudes and behaviours, which to some extent reflected the hodgepodge nature of classroom teachers' grading practices (Brookhart & Nitko, 2008).

The *How* component of the *Making-judgment* step was found only relevant to those recognized CA practices. The teachers were all concerned about the quality and fairness of their grading as well as the effect of their grading on students.

While the teachers' SA practices were mainly used for reporting purposes, their large number of FA practices were mainly used to promote student learning, especially through providing task-referenced feedback. Their dual-purpose CA practices were used not only for reporting but also for promoting learning and sometimes even improving teaching. Students in the three contexts all found their teachers' dual-purpose CA practices more beneficial and motivating than those SA and FA practices. This to some extent reflects the fact that grading is still a powerful motivator for Chinese students but to make it conducive to learning, students should be given the chances to learn and make improvement after the assessment practices.

In spite of the above-mentioned similarities, the three teachers' CA practices also varied along four dimensions. Concerning the methods of eliciting information about student learning during daily classroom teaching, the teachers varied from conducting more planned assessment-involving instructional activities/episodes (Andrew's case) to conducting more incidental assessment episodes (Mary's case), with Linda's case somewhere in between. Concerning the transparency of CA practices to students, the teachers varied from very transparent (Linda's case) to very vague (Mary's case), with Andrew's case somewhere in between. As to how teachers made their judgment, the

teachers varied from being more intuitive (Mary's case) to strictly following given criteria (Andrew's case), with Linda's case somewhere in between, since she both followed the generally agreed criteria and made localized adjustments so as not to de-motivate students. Regarding the relationship between assessment constructs and course learning objectives, the three cases varied from a high alignment (Linda's case) to a low alignment (Mary's case), with Andrew's case somewhere in between.

Taking each teacher's CA practices as a whole, the study revealed that Linda's FA and SA practices worked together like an upward spiral pushing students to achieve the course learning goals. In contrast, Andrew's SA and FA practices were like two parallel lines, and Mary's SA and FA practices were in partial synergy with the remaining CA practices spreading out like unchained pearls.

Such variations were found attributable to their specific teaching contexts, their teaching and/or assessment experience and expertise, as well as their beliefs about assessment or the best way to improve students' oral English.

8.2 Implications

This study has developed a descriptive tool to describe a teacher's CA practices (Figure 3.3), which can be applied in future studies to disentangle the complexity of teachers' actual CA practices.

As a classroom-based research situated in the EFL contexts in China, this study offers useful insights for EFL classroom teachers. This study demonstrated the usefulness of formative use of SA, as reflected through the dual-purpose CA practices. Probably this is the point where the Chinese context and western ideas of FA could be combined and work the best in the local context. Since Linda's case demonstrated the highest level synergy between FA and SA and her students found her CA practices very beneficial, probably some of her practices could offer some guidelines for other EFL teachers about how to conduct effective CA.

First, she acknowledged the power of scores for Chinese students and made the best use of them. She divided the total score into several parts and made sure that students

had to complete different tasks throughout the semester to obtain those scores rather than just take one test at the end. This could prevent students from only studying near the end after having idled through the whole semester. Also, the number of her marked assignments was limited to three and they were conducted during the latter half of the semester rather than spread out through the whole semester. Such practice was a modification of what she had done the previous year, when she gave students five marked assignments, one every three weeks, from the beginning of the semester (Linda_BI). With such changes, students would have time to accumulate their relevant knowledge and practice their skills before they were tested. In addition, she used scores to push students to do peer evaluation and self-evaluation. This could encourage them to learn from other students and reflect on their own performances.

Second, her CA practices were linked together by turning the course objectives into assessment constructs, displaying a chained spiral pattern. In this way, both teaching and learning were made goal-oriented, and any progress could be easily identified. Otherwise, if there were no such kind of link, both teaching and learning could become unfocused and students might not get a strong sense of achievement. This showed that both teachers and students need to be clear about their objectives and standards. In this way, teaching and learning can become more effective and efficient.

Third, though she was teaching a language skill course, she emphasized the development of students' higher-level thinking skills in this course. Taking this aspect as one of the course objectives, Linda emphasized critical thinking a lot by giving feedback on students' ideas and their thinking as well as giving them suggestions on how to make their ideas more logical and convincing. The fact that her students enjoyed such discussions and felt them useful suggests that the assessment in a language skill course should not be confined to the linguistic aspect, but should incorporate the cognitive aspect as well. A combination of the two may help students improve in both aspects.

Fourth, she encouraged students' involvement in the CA practices. As many researchers have pointed out the value of peer-assessment and self-assessment (e.g., Cheng & Warren, 2005; Falchikov & Boud, 1989; Little, 2011), Linda had found her

own way to incorporate such practices into her teaching, not only in working on the marked assignments but also in her daily classroom teaching. In this way, she made the CA practices transparent to students, pushed them to reflect on their performances, and at the same time encouraged them to assume more responsibility for their own learning.

In sum, it can be seen that Linda had used CA practices as tools to help with her teaching and students' learning, rather than just treating assessment as an end-of-semester event. It proved that CA, when used effectively, can be a powerful tool for classroom teaching and learning.

Moreover, as a descriptive study aiming at depicting and profiling the teachers' actual CA practices in their naturalistic contexts, this study carries important implications for EFL teachers' professional development in China. Although the three teachers were all from top universities in China and their teaching experiences ranged from 7 years to 22 years, none of them had received systematic training on CA, and the study findings also revealed their limited understanding of CA and their inadequacy in carrying out effective CA practices, especially in Andrew's and Mary's cases. This obviously calls for comprehensive and effective teacher training programs where teachers' awareness of CA should be enhanced, their knowledge about CA should be broadened, and their skills of carrying out CA should be developed. In such training programs, teachers should not only be provided with information about the purposes, constructs and procedures of CA but also be given opportunities to design, carry out and reflect on their own CA practices.

8.3 Limitations and suggestions

This study was intended to obtain an in-depth and contextualized understanding of the participant teachers' CA practices. While efforts were made to achieve such purposes, the study design was still limited in that the author could only visit each site four times within a one-semester course, which made the data collection like a series of snapshots of the teachers' actual CA practices. A more comprehensive and truer picture would have been obtained if the researcher could follow each teacher for a longer time and make more visits to each site.

The second limitation is the representativeness of the cases studied. The present study was based on convenience sampling, and the three teachers were all experienced oral English teachers from top universities in China. Therefore, more studies should be carried out in other language skill courses such as EFL reading, writing, and listening, and at different school levels such as the primary-school, middle-school, high-school levels, and at vocational schools. Even at the university level, teachers from different types of universities with various years of teaching experience should be recruited to participate. In this way, more representative and more comprehensive patterns of CA practices among Chinese teachers of English can be generated.

Third, it was not possible to pilot the student questionnaires as fully as would have been desirable and consequently they did not function as well as expected. This to some extent affected the quality of the data collected from the first two case studies. This highlights the need for careful design and piloting of questionnaires for use in this kind of study.

Fourth, the present study mainly relied on the self-report data of a limited number of volunteer students to capture the effects of CA. Future research should consider how, in a longitudinal study, to obtain data from every student in an efficient and effective way to reveal the effects of CA, especially those FA practices that are embedded within daily teaching, on students.

Finally, the present study was a description and analysis of existing practices. While the thick descriptions presented in this study might raise some teachers' awareness of CA and provide them with some guidelines as to how to conduct their own CA practices, more systematic studies are called for that would address how to raise teachers' awareness of CA and how to help teachers design and conduct effective CA practices in their own localized contexts.

References

Airasian, P. W., & Russell, M. K. (2008). *Classroom assessment: Concepts and applications* (sixth ed.). Boston: McGraw-Hill.

Alderson, J. C. (1986). Innovations in language testing. In M. Portal (Ed.), *Innovations in language testing: Proceedings of the IUS/NFER conference* (pp. 93-105). Windsor: NFER-Nelson.

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*(2), 115-129.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84, 261-271.

Anderson, R. S. (1998). Why talk about different ways to grade? The shift from traditional assessment to alternative assessment. *New Directions for Teaching and Learning, 74*, 5-16.

Andrade, H. L. (2010). Summing up and moving forward: Key challenges and future directions for research and development in formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 344-351). New York: Routledge.

Andrade, H. L., & Cizek, G. J. (Eds.). (2010). *Handbook of formative assessment*. New York: Routledge.

Angelo, T., A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers* (2nd ed.). San Francisco: Jossey-Bass Publishers.

ARG. (1999). Assessment for learning: Beyond the black box. Cambridge, UK: University of Cambridge School of Education.

ARG. (2002). *Assessment for learning: 10 principles*. University of Cambridge: Assessment Reform Group.

Bachman, L. F. (1990). *Fundamental considerations of language testing*. Oxford: Oxford University Press.

Bachman, L. F. (1991). What does language testing have to offer? *TESOL Quarterly, 25*, 671-747.

Bachman, L. F., & Palmer, A. (2010). *Language Assessment in Practice: Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.

Banks, S. R. (2005). *Classroom Assessment: Issues and Practices*. Boston: Pearson.

Bassey, M. (1999). *Case study research in educational settings*. Buckingham: Open University Press.

Bell, L. (1999). Public and private meaning in diaries: Researching family and children. In A. Bryman & R. G. Burgess (Eds.), *Qualitative Research* (Vol. 2). London: Sage.

Birjandi, P., & Tamjid, N. H. (2012). The role of self-, peer and teacher assessment in promoting Iranian EFL learners' writing performance *Assessment & Evaluation in Higher Education, 37*(5), 513-533. doi:10.1080/02602938.2010.549204

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

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through cassroom assessment. . *Phi Delta Kappan, 80*, 139-149.

Black, P., & Wiliam, D. (1998c). *Inside the black box: Raising standards through classroom assessment*. London: School of Education, King's College London.

Black, P., & Wiliam, D. (2003). 'In praise of educational research': Formative assessment. *British Educational Research Journal, 29*(5), 623-637.

Black, P., & Wiliam, D. (2006a). Assessment for learning in the classroom. In J. Gardner (Ed.), *Assessment and Learning* (pp. 9-25). London: SAGE Publications.

Black, P., & Wiliam, D. (2006b). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and Learning* (pp. 81-100). London: SAGE Publications.

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

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on the formative and summative evaluation of student learning*. New York: McGraw-Hill.

Booth, D. (2012). Scaffolding during the formal assessment of young EAL learners: A New Zealand case study. *Australian Review of Applied Linguistics, 35*(1), 5-27.

Brindley, G. (1998). Outcomes-based assessment and reporting in language learning programmes: A review of the issues. *Language Testing, 15*(1), 45-85. doi:10.1177/026553229801500103

Brindley, G. (2001). Outcomes-based assessment in practice: Some examples and emerging insights. *Language Testing, 18*(4), 393-407. doi:10.1177/026553220101800405

Brookhart, S. M. (1997). A theoretical framework for the role of classroom assessment in motivating student effort and achievement. *Applied measurement in Education, 10*(2), 161-180.

Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice, 22*(4), 5-12.

Brookhart, S. M. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teachers College Record, 106*(3), 429-458.

Brookhart, S. M. (2010). Mixing it up: Combining sources of classroom achievement information for formative and summative purposes. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 279-296). New York: Routledge.

Brookhart, S. M., & DeVoge, J. G. (1999). Testing a theory about the role of classroom assessment in student motivation and achievement. *Applied Measurement in Education, 12*, 409-425.

Brookhart, S. M., & DeVoge, J. G. (2000). *Classroom assessment, student motivation, and achievement in elementary and middle school*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Brookhart, S. M., & Durkin, D. T. (2003). Classroom assessment, student motivation, and achievement in high school social studies classes. *Applied Measurement in Education, 16*(1), 27-54.

Brookhart, S. M., & Nitko, A. J. (2008). *Assessment and grading in classrooms*. Upper Saddle River, New Jersey: Pearson Education, Inc.

Brookhart, S. M., Walsh, M., & Zientarski, W. A. (2006). The dynamics of motivation and effort for classroom assessments in middle school science and social studies. *Applied Measurement in Education, 19*(2), 151-184.

Bryman, A. (2001). *Social research methods*. Oxford: Oxford University Press.

Butler, S. M., & McMunn, N. D. (2006). *A teacher's guide to classroom assessment: Understanding and using assessment to improve student learning*. Hoboken, NJ, USA: Jossey-Bass.

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. doi:10.1002/j.1545-7249.2009.tb00243.x

Calfee, R. C., & Masuda, W. V. (1997). Classroom assessment as inquiry. In G. D. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement, and adjustment* (pp. 69-102). San Diego: Academic Press.

Cao, R., Zhang, W., & Zhou, Y. (2004). Implementation of formative assessment in an EFL writing course for Chinese non-English-major undergraduates 形成性评估在中国大学非英语专业英语写作教学中的运用. *Foreign Language Education, 25*(5), 82-87. doi:CNKI:SUN:TEAC.0.2004-05-020

Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International, 44*(1), 57-66. doi:10.1080/14703290601081332

Carless, D. (2008). Developing productive synergies between formative and summative assessment processes. In M. F. Hui & D. Grossman (Eds.), *Improving teacher education through action research* (pp. 9-23). New York: routledge.

Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings*. New York: Routledge.

Carless, D., Joughin, G., Liu, N.-F., & Associates. (2006). *How assessment supports learning: Learning-oriented assessment in action*. Hong Kong: Hong Kong University Press.

Carless, D., Joughin, G., & Mok, M. M. C. (2006). Learning-oriented assessment: Principles and practice. *Assessment & Evaluation in Higher Education, 31*(4), 395-398. doi:10.1080/02602930600679043

Celce-Murcia, M. (2008). Rethinking the role of communicative competence in language teaching. In E. A. Soler & P. S. Jorda (Eds.), *Intercultural Language Use and Language Learning* (pp. 41-57). Dordrecht, The Netherlands: Springer.

Cheng, L. (2005). *Changing language teaching through language testing*. Cambridge: Cambridge University Press.

Cheng, L. (2008). The key to success: English language testing in China. *Language Testing, 25*(1), 15-38. doi:10.1177/0265532207083743

Cheng, L., & Curtis, A. (2010a). The realities of English language assessment and the Chinese learner in China and beyond. In Cheng & A. Curtis (Eds.), *English language assessment and the Chinese learner* (pp. 3-12). New York: Routledge.

Cheng, L., & Curtis, A. (Eds.). (2010b). *English language assessment and the Chinese learner*. New York: Routledge.

Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods and procedures. *Language Testing, 21*(3), 360-389. doi:10.1191/0265532204lt288oa

Cheng, L., Rogers, T., & Wang, X. (2008). Assessment purposes and procedures in ESL/EFL classrooms. *Assessment & Evaluation in Higher Education, 33*(1), 9-32. doi:10.1080/02602930601122555

Cheng, L., & Wang, X. (2007). Grading, feedback, and reporting in ESL/EFL classrooms. *Language Assessment Quarterly, 4*(1), 85-107. doi:10.1080/15434300701348409

Cheng, L., Watanabe, Y. J., & Curtis, A. (Eds.). (2004). *Washback in language testing: Research contexts and methods*. London: Lawrence Erlbaum Associates.

Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language Testing, 22*(1), 93-121.

Cizek, G. J. (1997). Learning, achievement, and assessment: Constructs at a crossroads. In G. D. Phye (Ed.), *Handbook of classroom assessment* (pp. 1-32). NY: Academic Press.

Cizek, G. J. (2010). An introduction to formative assessment: History, characteristics, and challenges. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 3-17). New York: Routledge.

Clapham, C. (2000). Assessment and testing. *Annual Review of Applied Linguistics, 20*(1), 147-161. doi:10.1017/s0267190500200093

Colby-Kelly, C., & Turner, C. E. (2007). AfL research in the L2 classroom and evidence of usefulness: Taking formative assessment to the next level. *The Canadian Modern Language Review, 64*(1), 9–38. doi:10.1353/cml.2008.0019

Condon. W, and Hamp-Lyons, L. (1991) Introducing a Portfolio-Based Writing Assessment: Progress through Problems. In P. Belanoff & M. Dickson (Eds.), *Portfolios: Process and Product,* (pp. 231-247). Portsmouth: Boynton/Cook.

Cowie, B., & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education, 6*(1), 101-116.

Cronbach, L. (1964). Evaluation for course improvement. In R. Heath (Ed.), *New Curricula* (pp. 231-249). New York: Harper & Row.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*(4), 438-481.

Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong Secondary Schools. *Language Testing, 21*(3), 305-334.

Davison, C. (2007). Views from the chalkface: English language school-based assessment in Hong Kong. *Language Assessment Quarterly, 4*(1), 37-68.

Davison, C., & Cummins, J. (2007a). Assessment and evaluation in ELT: Shifting paradigms and practices. In J. Cummins & C. Davison (Eds.), *International Handbook of English Language Teaching* (Vol. 15, pp. 415-420): Springer US.

Davison, C., & Cummins, J. (2007b). Introduction: Assessment and Evaluation in ELT: Shifting Paradigms and Practices. In J. Cummins & C. Davison (Eds.), *International Handbook of English Language Teaching*, (Vol. 15, pp. 415-420): Springer US.

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly, 43*(3), 393-415. doi:10.1002/j.1545-7249.2009.tb00242.x

Dornyei, Z. (2003). *Questionnaires in second language research: Construction, administration, and processing*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Dornyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodoligies*. Oxford: Oxford University Press.

Douglas, D., & Chapelle, C. (Eds.). (1993). *A new decade of language testing research: Selected papers from the 1990 Language Testing Research Colloquium*. Alexandria, Virginia: Teachers of English to Speakers of Other Languages, Inc. .

Duff, P. A. (2008). *Case study research in applied linguistics*. New York: Taylor & Francis Group.

Elliott, J. (2003). Dynamic assessment in educational settings: Realising potential. *Educational Review, 55*(1), 15-32. doi:10.1080/0013191022000037830

Ellis, R. (2012). *Language teaching research and language pedagogy*. Malden, Mass: Wiley-Blackwell.

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.

Engeström, Y. (1987). *Learning by Expanding: An Activity Theoretical Approach to Developmental Research* Helsinki, Finland: Orienta-Konsultit.

Engeström, Y. (1999). Activity theory and individual social transformation. In Y. Engeström, R. Miettinen & R. L. Punamaki (Eds.), *Perspectives on Activity Theory* (pp. 19-38). New York: Cambridge University Press.

Engeström, Y. (2001). Expansive learning at work: Toward an activity theoretical reconceptualization. *Journal of Education and Work, 14*(1), 133-156. doi:10.1080/13639080020028747

English, F. W. (1992). *Deciding what to teach and test: Developing, aligning, and auditing the curriculum*. Newbury Park, California: Corwin Press, Inc.

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59*(4), 395-430.

Fontana, A., & Frey, J. (2003). The interview: From structured questions to negotiated text. In N. K. Denzin & Y. S. Lincoln (Eds.), *Collecting and interpreting qualitative materials* (2nd ed., pp. 61-106). Thousand Oaks, Calif.: Sage.

Fox, J. (2008). Alternative assessment. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (Vol. 7: Language testing and assessment, pp. 97-108). Online: Springer Science+Business Media LLC.

Fulcher, G. (2003). *Testing second language speaking*. London: Pearson Education Limited.

Gardner, J. (2006a). Assessment and learning: An introduction. In J. Gardner (Ed.), *Assessment and learning* (pp. 1-5). London: SAGE.

Gardner, J. (Ed.). (2006b). *Assessment and learning*. London: SAGE.

Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer Press.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist, 18*, 519-521.

Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education* Cambridge: Cambridge University Press.

Greenall, S., Tomalin, M., & Friedland, D. (Eds.). (2008). *New standard college English: Real communication listening and speaking 2*. Beijing: Foreign Language Teaching and Research Press.

Grigorenko, E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disabilities, 42*(2), 111-132. doi:10.1177/0022219408326207

Hamayan, E. V. (1995). Approaches to alternative assessment. *Annual Review of Applied Linguistics, 15*, 212-226. doi:10.1017/S0267190500002695

Hamp-Lyons, L. & Condon, W. (1993). Questioning assumptions about portfolio-based assessment. *College Composition and Communication, 44*(2), 176-190.

Han, M., & Yang, X. (2001). Educational assessment in China: Lessons from history and future prospects. *Assessment in Education, 8*(1), 5-10. doi: 10.1080/09695940120033216

Harlen, W. (2005). Teachers' summative practices and assessment for learning −Tensions and synergies. *The Curriculum Journal, 16*(2), 207-223.

Harlen, W. (2006). On the relationship between assessment for summative and formative purposes. In J. Gardner (Ed.), *Assessment and Learning* (pp. 103-117). London: Sage Publications.

Harlen, W. (2007). *Assessment of learning*. Los Angeles: SAGE.

Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education, 4*(3), 365-380.

Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Hill, K. (2012). *Classroom-based assessment in the school foreign language classroom*. Frankfurt am Main: Peter Lang.

Hill, K., & McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing, 29*(3), 395-420. doi: 10.1177/0265532211428317

Huang, S.-C. (2010). Convergent vs. divergent assessment: Impact on college EFL students' motivation and self-regulated learning strategies. *Language Testing, 28*(2), 251-271. doi:10.1177/0265532210392199

Huerta-Macias, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal, 5*, 8-11.

Hughes, R. (2011). *Teaching and researching speaking* (2nd ed.). London: Pearson Education Limited.

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing, 25*(3), 385-402. doi: 10.1177/0265532208090158

Ishihara, N. (2009). Teacher-based assessment for foreign language pragmatics. *TESOL Quarterly, 43*(3), 445-470. doi:10.1002/j.1545-7249.2009.tb00244.x

Jin, L., & Cortazzi, M. (2006). Changing practices in Chinese cultures of learning. *Language, Culture, and Curriculum, 19*(1), 5-20. doi: 10.1080/07908310608668751

Keppell, M., & Carless, D. (2006). Learning-oriented assessment: A technology-based case study. *Assessment in Education, 13*(2), 179-191. doi: 10.1080/09695940600703944

Lantolf, J. P. (2006). Sociolcultural theory and L2: State of the art. *Studies in Second Language Acquisition, 28*(1), 67-109. doi:10.1017/s0272263106060037

Lantolf, J. P. (Ed.). (2000). *Sociocultural theory and second language learning*. Oxford: Oxford University Press.

Law, B., & Eckes, M. (1995). *Assessment and ESL: On the yellow big road to the withered of Oz*. Winnipeg, Manitoba, Canada: Peguis Publishers.

Leung, C. (2004). Developing formative teacher assessment: Knowledge, practice, and change. *Language Assessment Quarterly: An International Journal, 1*(1), 19-41.

Leung, C. (2005). Classroom teacher assessment of second language development: Construct as practice. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 869-888). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Leung, C. (2007). Dynamic assessment: Assessment for and as teaching. *Language Assessment Quarterly, 4*(3), 257-278. doi:10.1080/15434300701481127

Leung, C., & Mohan, B. (2004). Teacher formative assessment and talk in classroom contexts: Assessment as discourse and assessment of discourse. *Language Testing, 21*(3), 335-359. doi:10.1191/0265532204lt287oa

Lightbown, P., & Spada, N. (2006). *How languages are learned*. Oxford: Oxford University Press.

Little, D. (2011). Learner autonomy, self-assessment and language tests: Towards a new assessment culture. In B. Morrison (Ed.), *Independent Language Learning* (pp. 25-39). Hong Kong: Hong Kong University Press.

Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.

Lyle, J. (2003). Stimulated Recall: A report on its use in naturalistic research. *British Educational Research Journal, 29*(6), 861-878. doi:10.1080/0141192032000137349

Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, N. J.: Lawrence Erlbaum.

Maehr, M. L., & Meyer, H. A. (1997). Understanding motivation and school: Where we've been, where we are, and where we need to go. *Educational Psychology Review, 9*, 371-401.

Marzano, R. J., Pickering, D., & McTighe, J. (1993). *Assessing student outcomes: Performance assessment using the dimensions of learning model*. Alexandria, Virginia: Association for Supervision and Curriculum Development.

McKay, S. L. (2006). *Researching second language classrooms*. Mahwah, N.J.: Lawrence Erlbaum.

McMillan, J. H. (2007a). *Classroom assessment: Principles and practices for effective standards-based instruction* (4th ed.). Boston: Pearson Education, Inc.

McMillan, J. H. (2007b). Formative classroom assessment: The key to improving student achievement. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 1-7). New York: Teachers College, Columbia University.

McMillan, J. H. (Ed.). (2007c). *Formative classroom assessment: Theory into practice*. New York: Teachers College, Columbia University.

McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing, 18*(4), 333-349.

Merriam, S. B. (1988). *Case study research in education : A qualitative approach*. San Francisco: Jossey-Bass.

Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco: Jossey-Bass.

Messick, S. (1998). Test validity: A matter of consequency. *Social Indicators Research, 45*(1-3), 35-44. doi:10.1023/A:1006964925094

Middleton, M., & Midgley, C. (1997). Avoiding the demonstration of lack of ability: An underexplored aspect of goal theory. *Journal of Educational Psychology, 89*, 710-718.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded source-book* (2nd ed.). London: Sage.

Mitchell, R. (1992). *Testing for learning: How new approaches to evaluation can improve American schools*. New York: Free Press.

Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist, 22*(2), 155-175.

Neuman, W. L. (2003). *Social research methods: Qualitative and quantitative approaches* (5th ed.). Boston: Pearson Education, Inc.

Nunan, D., & Bailey, K. M. (2010). *Exploring second language classroom research: A comprehensive guide*. Beijing: Foreign Laguage Teaching and Research Press, authorized by Gengage Learning.

O'Malley, J. M., & Pierce, L. V. (1996). *Authentic assessment for English language learners: Practical approaches for teachers*. New York: Addison-Wesley.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks: SAGE.

Petrovsky, A. V. (1985). *Studies in psychology: The collective and the individual*. Moscow: Progress.

Phye, G. D. (Ed.). (1997). *Handbook of classroom assessment: Learning, achievement, and adjustment*. San Diego: Academic Press.

Poehner, M. E. (2007). Beyond the test: L2 dynamic assessment and the transcendence of mediated learning. *Modern Language Journal, 91*, 323-340.

Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development*. Berline: Springer.

Poehner, M. E. (2009). Group dynamic assessment: Mediation for the L2 classroom. *TESOL Quarterly, 43*(3), 471-491. doi:10.1002/j.1545-7249.2009.tb00245.x

Poehner, M. E., & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research, 9*(3), 233-265.

Popham, W. J. (1994). The instructional consequences of criterion-referenced clarity. *Educational Measurement: Issues and Practice, 13*(4), 15-18, 30.

Popham, W. J. (2008a). Classroom assessment: Staying instructionally afloat in an ocean of accountability. In C. A. Dwyer (Ed.), *The Future of Assessment: Shaping Teaching and Learning* (pp. 263-278). New York: Taylor & Francis Group.

Popham, W. J. (2008b). *Transformative assessment*. Alexandria, Virginia USA: Association for Supervision and Curriculum Development.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice, 48*(1), 4-11. doi:10.1080/00405840802577536

Popham, W. J. (2011). *Classroom assessment: What teachers need to know* (6th ed.). Boston: Pearson.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science, 28*(1), 4-13.

Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing, 18*(4), 429-462. doi:10.1177/026553220101800407

Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing, 21*(3), 249-258.

Rea-Dickins, P. (2006). Currents and eddis in the discourse of assessment: A learning-focused interpretation. *International Journal of Applied Linguistics, 16*(2), 163-188.

Rea-Dickins, P. (2007a). Classroom-based assessment: Possibilities and pitfalls. In J. Cummins & C. Davison (Eds.), *International Handbook of English Language Teaching* (Vol. 15, pp. 505-520): Springer US.

Rea-Dickins, P. (2007b). Learning or measuring? Exploring teacher decision-making in planning for classroom-based language assessment. In S. Fotos & H. Nassaji (Eds.), *Form-focused instruction and teacher education* (pp. 193-210). Oxford: Oxford University Press.

Rea-Dickins, P. (2008). Classroom-based language assessment. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of Language and Education* (2nd ed., Vol. 7: Language testing and assessment, pp. 257-271). Online: Springer Science+Business Media LLC.

Rea-Dickins, P., & Gardner, S. (2000). Snares and silver bullets: Disentangling the construct of formative assessment. *Language Testing, 17*(2), 215-243.

Rhodes, N. C., Rosenbusch, M., H., & Thompson, L. (1997). Foreign languages: Instruments, techniques, and standards. In G. D. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement, and adjustment* (pp. 381-415). San Diego: Academics Press.

Richards, L., & Morse, J. M. (2007). *Readme first for a user's guide to qualitative methods* (2nd ed.). London: Sage.

Ritchie, J., & Spencer, L. (2002). Qualitative data analysis for applied policy research. In A. M. Huberman & M. B. Miles (Eds.), *The Qualitative Researcher's Companion* (pp. 305-330). Thousand Oaks: Sage.

Ross, S. J. (2005). The impact of assessment method on foreign language proficiency growth. *Applied Linguistics, 26*(3), 317-342. doi:10.1093/applin/ami011

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119-144.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné & M. Scriven (Eds.), *Perspectives on curriculum evaluation* (pp. 39-83). Chicago, IL: Rand McNally.

Sheen, Y., & Ellis, R. (2011). Corrective feedback in language teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 593-610). New York: Routledge.

Shepard, L. A. (1989). Why we need better assessments. *Educational Researcher, 46*(7), 4-9.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4-14. doi:10.3102/0013189X029007004

Shepard, L. A. (2005). Linking formative assessment to scaffolding. *Educational Leadership, 63*(3), 66-70.

Shepard, L. A. (2006). Classroom assessment. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 624-646). Westport, CT: Praeger.

Shepard, L. A. (2008). Formative assessment: Caveat emptor. In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 279-303). New York: Erlbaum.

Simons, H. (2009). *Case study research in practice*. Los Angeles: Sage.

Sinclair, J. M., & Coulthard, M. (1975). *Towards an analysis of discourse*. London: Oxford University Press.

Spada, N., & Fröhlich, M. (1995). *COLT Communicative orientation of language teaching observation scheme : coding conventions and applications*. Sydney: National Centre for English Language Teaching and Research.

Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing. The nature and measurement of learning potential*. Cambridge: Cambridge University Press.

Stiggins, R. J. (2008). *An introduction to student-involved assessment for learning* (fifth ed.). Upper Saddle River, New Jersey: Pearson Education, Inc.

Stiggins, R. J., & Conklin, N. F. (1992). *In teacher's hands: Investigating the practices of classroom assessment*. Albany, NY: State University of New York.

Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. New York: Routledge.

Taras, M. (2005). Assessment--Summative and formative--Some theoretical reflections. *British Journal of Educational Studies, 53*(4), 466-478. doi:10.1111/j.1467-8527.2005.00307.x

Torrance, H. (1993). Formative assessment: Some theoretical problems and empirical questions. *Cambridge Journal of Education, 23*(3), 333-343.

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. Maidenhead: Open University Press.

Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal, 27*(5), 615-631. doi:10.1080/01411920120095780

van Lier, L. (1989). Reeling, writhing, drawling, stretching, and fainting: Oral proficiency interviews as conversation. *TESOL Quarterly, 23*(3), 489-508.

Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory* Cambridge: Cambridge University Press.

Wang, X., & Cheng, L. (2010). Chinese EFL students' perceptions of the classroom assessment environment and their goal orientations. In Cheng & A. Curtis (Eds.), *English Language Assessment and the Chinese Learner* (pp. 202-218). New York: Routledge.

Wette, R., & Barkhuizen, G. (2009). Teaching the book and educating the person: Challenges for university English language teachers in China. *Asia Pacific Journal of Education, 29*(2), 195-212. doi:10.1080/02188790902857180

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan, 70*(9), 703-713.

Wiliam, D. (2009). Assessment for learning: Why, what and how? London: Institute of Education, University of London.

Wiliam, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 18-40). New York: Routledge.

Wiliam, D., & Leahy, S. (2007). A theoretical foundation for formative assessment. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 29-42). New York: Teachers College, Columbia University.

Wiliam, D., & Thompson, M. (2008). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53-84). New York: Erlbaum.

Wood, D. (2010). *Formulaic language and second language speech fluency: Blackground, evidence and classroom applications*. London: Continuum International Publishing.

Xi, X., Higgins, D., Cechner, K., & Williamson, D. M. (2008). Automated scoring of spontaneous speech using SpeechRAter v1.0. ETS Research Report. http://www.ets.org/research/policy_research_reports/publications/report/2008/hukv.

Yin, M. (2010). Understanding classroom language assessment through teacher thinking research. *Language Assessment Quarterly, 7*(2), 175-194. doi:10.1080/15434300903447736

Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). Los Angeles: SAGE.

Zhou, P., & Qin, X. (2005). Applying formative assessment in the Internet teaching of College English 形成性评估在大学英语网络教学中的应用. *Foreign Language Audio-Visual Instruction 外语电化教学*(105), 9-13.

Zuengler, J., Ford, C., & Fassnacht, C. (1998). Analyst eyes and camera eyes: Theoretical and technological considerations in "seeing" the details of classroom interaction. Albany, N.Y.: School of Education, University at Albany.

Appendixes

Appendix 1. Classroom Observation Scheme

Teacher:
Date:

Time	Teacher's words and behavior	Students' words and behavior	Comments & reflections

Appendix 2. Teacher Interview Guide for Baseline Information

Demographic information

Age group
Gender
Highest degree
Research interest

Past teaching and learning experiences

Years of teaching English:
Years of teaching oral English course:
Received any teacher training? When? How long? About what?

Present teaching context

Information about the university
Information about the department
Information about the oral English course
Course objectives
Usual ways of conducting the oral English course
Usual ways of assessing students

Teacher belief about : Teaching, learning, and assessment

Arrangement for classroom observation

Appendix 3. Teacher End-of-Semester Interview Guide

General feelings about this course

Overall feeling/ opinion of your teaching over the whole semester and why
Students' achievement of the course objectives and contributing factors
Successful / unsuccessful teaching experiences: describe and explain

Teacher belief: Assessment, teaching, and learning

Classroom assessment practices

Summative assessment practices:

Final test: its functions, format, content, marking criteria, grading process
How to work out the final score for each student for this course
Any other forms of assessment: why, when, how, weightings

Formative assessment practices:

Any incidental assessments: describe and explain
About students' strengths and weaknesses: what has been found? How? For any use?
About students' effort: what has been found? How? For any use?
About students' anxiety: what has been found? How? For any use?
About students' learner autonomy: what has been found? How? For any use?
About your own instruction: general opinion, any modifications, when and why

Difficulties and expectations:

Difficulties you encountered when assessing your students
Help needed to get to improve your classroom assessment practices

Appendix 4. Teacher Stimulated Retrospective Interview Guide

Instruction to teachers before the interview:
"The purpose of this interview is to find out why you have conducted some of your classroom activities the way you have done them. I will remind you of those activities one by one by giving you a short summary, and after each reminder I will ask you a list of why and how questions concerning that activity. If you cannot remember that activity very clearly, I will replay the recording of that part for you."
Generic questions concerning each activity where assessment might be involved:
Why did you organize (a particular assessment activity)?
Why did you ask this question?
How did you evaluate this student's performance on this task?
Why did you (give this kind of feedback)? (The actual type of feedback will be used in this question)

Appendix 5a. Student Beginning-of-Semester Questionnaire (The Translated English Version)

Questionnaire on Oral English Learning (at the beginning of the semester)

Tracking code: _____

Part I: Some information about you

1 Gender: ☐ Male ☐ Female
2 Taken any English proficiency tests (e.g., TOEFL, IELTS, CET4, CET6, etc.): ☐ Yes ☐ No
If yes, the most recent time when you took such a test: _____,
name of the test _____ your score: _____
3 Type of middle school: ☐ Foreign language middle school ☐ Non-foreign-language middle school
4 English-learning starting age: ☐ before 8 ☐ 9-12 ☐ 13-15 ☐ after 16
5 Overseas experience: ☐ Yes ☐ No
   If yes, which country: _____ for how long: _____

Part II. Your oral English learning

Please circle the number that suits you the best.	Not true at all	Not really true	Partly true partly untrue	Mostly true	Absolutely true
1 I am very confident when I speak English.	1	2	3	4	5
2 I worry that native English speakers will find my oral English strange.	1	2	3	4	5
3 My oral English is very fluent.	1	2	3	4	5
4 My poor oral English makes me feel inferior to those who can speak fluent oral English.	1	2	3	4	5
5 When I speak in English, I can always find the words to express my ideas.	1	2	3	4	5
6 When I speak in English, I can say exactly what I want to say.	1	2	3	4	5
7 I am satisfied with my present oral English proficiency.	1	2	3	4	5
8 It's very important for me to completely grasp the skills taught in this course.	1	2	3	4	5
9 It is very important for me to be the top student in this oral English class.	1	2	3	4	5
10 I will be satisfied so long as I can pass this course.	1	2	3	4	5
11 I am eager to become a smart student in my teacher's eye in this course.	1	2	3	4	5
12 I am willing to spend a lot of time	1	2	3	4	5

practicing to improve my oral English.					
13 I worry that I may fail this course.	1	2	3	4	5
14 I wish to get a high score in this course.	1	2	3	4	5
15 I believe that if I work hard, my oral English will improve.	1	2	3	4	5
16 I feel that no matter how hard I try, my oral English will make little improvement.	1	2	3	4	5
17 I get very nervous when a foreigner asks me something in English.	1	2	3	4	5
18 I feel distressed about being unable to improve my oral English.	1	2	3	4	5
19 I worry that other students will laugh at me when I speak in English.	1	2	3	4	5
20 I actively create opportunities to talk with others in English.	1	2	3	4	5
21 I regularly enlarge my oral English vocabulary.	1	2	3	4	5
22 I try my best to improve the grammatical accuracy of my oral English.	1	2	3	4	5
23 I grasp every opportunity to practice my oral English.	1	2	3	4	5

This is the end of the questionnaire. Thank you very much!

Appendix 5b. Student Beginning-of-Semester Questionnaire (The Applied Chinese Version)

英语口语学习调查问卷 (学期初)

跟踪编码：_____
第一部分：您的个人信息
性别：   □ 男   □ 女
您是否参加过大规模的英语水平考试（如托福、雅思、四级、六级考试等）：   □ 是
□ 否
如果参加过，请写出最近一次考试的时间_____ 名称 _____分数: _____
所来自的中学：   □ 外语学校   □ 非外语学校
开始学英语的年龄： □ 8岁以下   □ 9-12岁   □ 13-15岁   □ 16岁以上
您是否有过出国经历：   □ 有   □ 没有
如果有，请写出您去的国家: _____   在那里停留的时间: _____

第二部分：您目前的英语口语学习状况

请根据实际情况在相应的数字上画圈。	根本不真实	基本不真实	部分真实	基本真实	绝对真实
1 我现在讲英语时非常自信。	1	2	3	4	5
2 我担心英语是母语的人会觉得我说的英语很奇怪。	1	2	3	4	5

	1	2	3	4	5
3 我目前的英语口语非常流利。	1	2	3	4	5
4 英语口语差使我觉得跟那些口语好的比起来，矮人一截。	1	2	3	4	5
5 我讲英语时总能找到词语来表达我的想法。	1	2	3	4	5
6 我讲英语时能讲得很准确。	1	2	3	4	5
7 我对我现在的英语口语水平很满意。	1	2	3	4	5
8 尽可能全面地掌握这门口语课上教的技能对我很重要。	1	2	3	4	5
9 在这个课上成为班级拔尖的学生对我很重要。	1	2	3	4	5
10 在这个课上只要考试能通过我就很满意。	1	2	3	4	5
11 我渴望成为老师眼中优秀的学生。	1	2	3	4	5
12 我愿意多花时间反复练习以提高自己的英语口语水平。	1	2	3	4	5
13 我很担心自己这门课会通不过。	1	2	3	4	5
14 我渴望在这门课上得高分。	1	2	3	4	5
15 我相信只要我努力我的英语口语一定会进步。	1	2	3	4	5
16 我觉得我的英语口语怎么练都不可能有大的提高。	1	2	3	4	5
17 如果一个外国人用英语问我什么事情，我会很紧张。	1	2	3	4	5
18 我的英语口语能力总也上不去，感到很苦恼。	1	2	3	4	5
19 我担心我讲英语的时候其他同学会笑话我。	1	2	3	4	5
20 我主动创造机会用英语会话。	1	2	3	4	5
21 我有规律地扩展自己的英语口语的词汇量。	1	2	3	4	5
22 我努力改进自己英语口语的语法准确性。	1	2	3	4	5
23 我抓住一切机会练习英语口语。	1	2	3	4	5

问卷到此结束！非常感谢！

Appendix 6a. Student End-of-Semester Questionnaire (The Translated English Version)

Questionnaire on Oral English Learning (at the end of the semester)

Tracking code: _____
Part I: Some information about you
1 **Gender:** ☐ Male   ☐ Female
2 **Taken any English proficiency tests** (e.g., TOEFL, IELTS, CET4, CET6, etc.):
            ☐ Yes   ☐ No
If yes, the most recent time when you took such a test: _____,
name of the test _____   your score: _____
3 **Type of middle school**: ☐ Foreign language middle school
                         ☐ Non-foreign-language middle school
4 **English-learning starting age:**   ☐ before 8   ☐ 9-12   ☐ 13-15   ☐ after 16
5 Overseas experience: ☐ Yes   ☐ No
    If yes, which country: _____   for how long: _____

Part II. Your oral English learning

Please circle the number that suits you the best.	Not true at all	Not really true	Partly true partly untrue	Mostly true	Absolutely true
1 I am very confident when I speak English.	1	2	3	4	5
2 I worry that native English speakers will find my oral English strange.	1	2	3	4	5
3 My oral English is very fluent.	1	2	3	4	5
4 My poor oral English makes me feel inferior to those who can speak fluent oral English.	1	2	3	4	5
5 When I speak in English, I can always find the words to express my ideas.	1	2	3	4	5
6 When I speak in English, I can say exactly what I want to say.	1	2	3	4	5
7 I am satisfied with my present oral English proficiency.	1	2	3	4	5
8 It's very important for me to completely grasp the skills taught in this course.	1	2	3	4	5
9 It is very important for me to be the top student in this oral English class.	1	2	3	4	5
10 I will be satisfied so long as I can pass this course.	1	2	3	4	5
11 I am eager to become a smart student in my teacher's eye in this course.	1	2	3	4	5
12 I am willing to spend a lot of time practicing to improve my oral English.	1	2	3	4	5
13 I worry that I may fail this course.	1	2	3	4	5
14 I wish to get a high score in this course.	1	2	3	4	5
15 I believe that if I work hard, my oral English will improve.	1	2	3	4	5
16 I feel that no matter how hard I try, my oral English will make little improvement.	1	2	3	4	5
17 I get very nervous when a foreigner asks me something in English.	1	2	3	4	5
18 I feel distressed about being unable to improve my oral English.	1	2	3	4	5
19 I worry that other students will laugh at me when I speak in English.	1	2	3	4	5
20 I actively create opportunities to talk with others in English.	1	2	3	4	5
21 I regularly enlarge my oral English vocabulary.	1	2	3	4	5
22 I try my best to improve the grammatical accuracy of my oral English.	1	2	3	4	5
23 I grasp every opportunity to practice my oral English.	1	2	3	4	5

Part III. Your perceptions of this oral English course

A. Classroom practices. Please circle the practices that your teacher has used in this oral English course, and then circle the numbers that best represent the frequency and usefulness of each selected practice.

Please circle the practices that your teacher has used in this oral English course.		How frequent?					How useful?				
		Once only	Every month	Every unit	Every lesson	Every activity	not at all	a little	somewhat	fairly	extremely
1	Explain the learning objectives of the whole course.	1	2	3	4	5	1	2	3	4	5
2	Explain the learning objectives of a particular unit.	1	2	3	4	5	1	2	3	4	5
3	Explain the learning objectives of a particular activity.	1	2	3	4	5	1	2	3	4	5
4	Explain the connections between the learning objectives of the whole course, of a particular unit, and of a particular activity.	1	2	3	4	5	1	2	3	4	5
5	Explain the skills to be grasped in this course.	1	2	3	4	5	1	2	3	4	5
6	Organize activities for students to practice the skills to be grasped in this course.	1	2	3	4	5	1	2	3	4	5
7	Observe how well students have grasped the skills.	1	2	3	4	5	1	2	3	4	5
8	Use questioning to check how well students have grasped the skills.	1	2	3	4	5	1	2	3	4	5
9	Comment on students' performances after they finish a classroom activity.	1	2	3	4	5	1	2	3	4	5
10	Point out students' strengths and weaknesses after they finish a classroom activity.	1	2	3	4	5	1	2	3	4	5
11	Tell students the correct answers after students finish a classroom activity.	1	2	3	4	5	1	2	3	4	5
12	Tell students how to do better next time after students finish a classroom activity.	1	2	3	4	5	1	2	3	4	5
13	Praise those students who perform well in doing classroom activities.	1	2	3	4	5	1	2	3	4	5
14	Criticize those students who do not try their best to do a classroom activity.	1	2	3	4	5	1	2	3	4	5
15	Encourage students not to be afraid of making mistakes.	1	2	3	4	5	1	2	3	4	5
16	Organize students to do self-assessment.	1	2	3	4	5	1	2	3	4	5
17	Organize students to do peer-assessment.	1	2	3	4	5	1	2	3	4	5
18	Assign after-class homework.	1	2	3	4	5	1	2	3	4	5
19	Check students' homework in class.	1	2	3	4	5	1	2	3	4	5
20	Explain to the students how they are assessed in this course.	1	2	3	4	5	1	2	3	4	5
21	Inform students what they are expected to do at the course test(s).	1	2	3	4	5	1	2	3	4	5
22	Explain to students how they can get high scores in the course test(s).	1	2	3	4	5	1	2	3	4	5
23	If there are other practices that are not included in the above list, please write them down here, and circle the best numbers for										

		1	2	3	4	5	1	2	3	4	5
each practice.		1	2	3	4	5	1	2	3	4	5
		1	2	3	4	5	1	2	3	4	5
		1	2	3	4	5	1	2	3	4	5

B. Overall impression. If you have other comments or thoughts about this oral English course, please write them down here.

This is the end of the questionnaire. Thank you very much!

Appendix 6b. Student End-of-Semester Questionnaire (The Applied Chinese Version)

英语口语学习调查问卷 (学期末)

跟踪编码：_____

第一部分：您的个人信息
性别:　　□ 男　　　□ 女
您是否参加过大规模的英语水平考试（如托福、雅思、四级、六级考试等）：　　□ 是 □ 否
如果参加过，请写出最近一次考试的时间_____名称 _____分数: _____
所来自的中学:　　□ 外语学校　　　　□ 非外语学校
开始学英语的年龄: □ 8岁以下　□ 9-12岁　　□ 13-15岁　　□ 16岁以上
您是否有过出国经历:　□ 有　　　□ 没有
如果有，请写出您去的国家: _____　在那里停留的时间: _____

第二部分：您目前的英语口语学习状况

请根据实际情况在相应的数字上画圈。	根本不真实	基本不真实	部 分真实	基 本真实	绝 对真实
1　我现在讲英语时非常自信。	1	2	3	4	5
2　我担心英语是母语的人会觉得我说的英语很奇怪。	1	2	3	4	5
3　我目前的英语口语非常流利。	1	2	3	4	5
4　英语口语差使我觉得跟那些口语好的比起来，矮人一截。	1	2	3	4	5
5　我讲英语时总能找到词语来表达我的想法。	1	2	3	4	5
6　我讲英语时能讲得很准确。	1	2	3	4	5
7　我对我现在的英语口语水平很满意。	1	2	3	4	5
8　尽可能全面地掌握这门口语课上教的技能对我很重要。	1	2	3	4	5
9　在这个课上成为班级拔尖的学生对我很重要。	1	2	3	4	5
10　在这个课上只要考试能通过我就很满意。	1	2	3	4	5
11　我渴望成为老师眼中优秀的学生。	1	2	3	4	5
12　我愿意多花时间反复练习以提高自己的英语	1	2	3	4	5

口语水平。					
13　我很担心自己这门课会通不过。	1	2	3	4	5
14　我渴望在这门课上得高分。	1	2	3	4	5
15　我相信只要我努力我的英语口语一定会进步。	1	2	3	4	5
16　我觉得我的英语口语怎么练都不可能有大的提高。	1	2	3	4	5
17　如果一个外国人用英语问我什么事情，我会很紧张。	1	2	3	4	5
18　我的英语口语能力总也上不去，感到很苦恼。	1	2	3	4	5
19　我担心我讲英语的时候其他同学会笑话我。	1	2	3	4	5
20　我主动创造机会用英语会话。	1	2	3	4	5
21　我有规律地扩展自己的英语口语的词汇量。	1	2	3	4	5
22　我努力改进自己英语口语的语法准确性。	1	2	3	4	5
23　我抓住一切机会练习英语口语。	1	2	3	4	5

第三部分：对口语课的感受

A. 课堂做法：请依据实际情况圈出在这门课上出现的课堂做法、及其出现的频率和对你学习的帮助大小。

请圈出老师在这门课上使用的课堂做法。	出现的频率					对自己学习的帮助				
	整个学期一次	每月	每个单元	每次课	每个活动	根本没用	不太有用	有一些用	用处比较大	非常有用
1 阐明这门课的学习目标。	1	2	3	4	5	1	2	3	4	5
2 阐明一个单元的学习目标。	1	2	3	4	5	1	2	3	4	5
3 阐明一个活动的学习目标。	1	2	3	4	5	1	2	3	4	5
4 阐明不同层次的学习目标之间的关系。	1	2	3	4	5	1	2	3	4	5
5 阐明学生需要掌握的技能。	1	2	3	4	5	1	2	3	4	5
6 组织活动让学生练习需要掌握的技能。	///	2	3	4	5	1	2	3	4	5
7 通过观察来检查学生对所学技能的掌握情况。	1	2	3	4	5	1	2	3	4	5
8 通过提问来检查学生对所学技能的掌握情况。	1	2	3	4	5	1	2	3	4	5
9 评价学生在练习	1	2	3	4	5	1	2	3	4	5

活动中的表现。										
10 指明学生在练习活动中表现出的优缺点。	1	2	3	4	5	1	2	3	4	5
11 练习活动后告诉学生正确答案。	1	2	3	4	5	1	2	3	4	5
12 练习活动后告诉学生如何进步。	1	2	3	4	5	1	2	3	4	5
13 表扬课堂上表现好的学生。	1	2	3	4	5	1	2	3	4	5
14 批评课堂上不认真的学生。	1	2	3	4	5	1	2	3	4	5
15 鼓励学生不要怕犯错误。	1	2	3	4	5	1	2	3	4	5
16 组织学生进行自我评价。	1	2	3	4	5	1	2	3	4	5
17 组织学生与同伴进行互相评价。	1	2	3	4	5	1	2	3	4	5
18 布置课外作业。	1	2	3	4	5	1	2	3	4	5
19 课上检查学生的课外作业。	1	2	3	4	5	1	2	3	4	5
20 阐明这门课的考核方式。	1	2	3	4	5	1	2	3	4	5
21 告诉学生在考试中应该如何表现。	1	2	3	4	5	1	2	3	4	5
22 向学生指出如何在考试中得高分。	1	2	3	4	5	1	2	3	4	5
23 如果还有其它做法，请写在下面：	1	2	3	4	5	1	2	3	4	5
	1	2	3	4	5	1	2	3	4	5
	1	2	3	4	5	1	2	3	4	5

B. 整体感受：关于这门课您如果还有其它方面的感受，请写在下面。

问卷到此结束！非常感谢！

Appendix 7a. Student End-of-Semester Questionnaire (The Translated English Version)

Questionnaire on Oral English Learning (at the end of the semester)

Tracking code: _____

Part I: Some information about you
Gender: ☐ Male ☐ Female
Taken any English proficiency tests (e.g., TOEFL, IELTS, CET4, CET6, etc.): ☐ Yes ☐ No
If yes, the most recent time when you took such a test: _____, name of the test
_____ your score: _____
Type of middle school:
☐ Foreign language middle school ☐ Non-foreign-language middle school
English-learning starting age: ☐ before 8 ☐ 9-12 ☐ 13-15 ☐ after 16
Overseas experience: ☐ Yes ☐ No
    If yes, which country: _____ for how long: _____

Part II. In this course, when you were involved in the following activities, did you think your teacher was assessing you? Would your performances in those activities affect your final scores in this course? Please circle your answers according to the real situations.

When you were involved in the following activity:	Was your teacher assessing you?		Would your performance affect your final score?	
1 Arriving at the classroom on time	Yes	No	Yes	No
2 Giving a news report at the beginning of a class	Yes	No	Yes	No
3 Answering your teacher's questions raised during her lecture	Yes	No	Yes	No
4 Presenting your ideas during a group activity while your teacher was around listening	Yes	No	Yes	No
5 Presenting your ideas before the whole class after a group activity	Yes	No	Yes	No
6 After giving a speech, responding to the questions from your classmates and your teacher	Yes	No	Yes	No
7 Giving three speeches before the whole classes as required	Yes	No	Yes	No
8 Giving comments about your classmates' performances	Yes	No	Yes	No
9 Reflecting on your own strengths and weaknesses in giving a speech	Yes	No	Yes	No

Part III. Your sense of achievement

| Compared with the beginning of the semester, in terms of | yours has become / is ____. | | | | |
	much better / stronger (5)	a little better / stronger (4)	just the same (3)	a little worse / weaker (2)	much worse / weaker (1)
1 pronunciation and intonation when speaking English	(5)	(4)	(3)	(2)	(1)
2 fluency when speaking English	(5)	(4)	(3)	(2)	(1)
3 vocabulary size of my oral English	(5)	(4)	(3)	(2)	(1)
4 accuracy of my oral English	(5)	(4)	(3)	(2)	(1)
5 complexity of the ideas that I can express in English orally	(5)	(4)	(3)	(2)	(1)
6 communicative ability when I talk with others in English	(5)	(4)	(3)	(2)	(1)

7 understanding the skills and techniques of giving a public speech in English	(5)	(4)	(3)	(2)	(1)
8 understanding how to improve my public speaking ability step by step	(5)	(4)	(3)	(2)	(1)
9 my ability to give a public speech	(5)	(4)	(3)	(2)	(1)
10 my ability to debate in English	(5)	(4)	(3)	(2)	(1)
11 the belief that I will be able to speak excellent English in future	(5)	(4)	(3)	(2)	(1)
12 the belief that I will become a very good public speaker in future	(5)	(4)	(3)	(2)	(1)
13 the belief that I will make progress so long as I work hard	(5)	(4)	(3)	(2)	(1)
14 the belief that I will not improve no matter how hard I try	(5)	(4)	(3)	(2)	(1)
15 the confidence in giving a public speech in English	(5)	(4)	(3)	(2)	(1)
16 the anxiety caused by giving a public speech in English	(5)	(4)	(3)	(2)	(1)
17 the hope to get a good score in this course	(5)	(4)	(3)	(2)	(1)
18 the hope to be a good student in the teacher's eyes	(5)	(4)	(3)	(2)	(1)
19 the worry to get a poor score in this course	(5)	(4)	(3)	(2)	(1)
20 the effort to create opportunities to practice my public speaking skills	(5)	(4)	(3)	(2)	(1)
21 the effort to enlarge my oral English vocabulary	(5)	(4)	(3)	(2)	(1)
22 the effort to improve the accuracy of my oral English	(5)	(4)	(3)	(2)	(1)

Part IV. Which of the following practices have/has brought about your achievement?

Please circle the practices that your teacher has used in this oral English course.	In what aspect(s) is a selected practice helpful to your study? (please circle) To what degree is the practice useful to the aspect(s) selected? (please link)					
1 Explain the learning	aspect(s)	Develop knowledge	Improve skill	Boost confidence	Enhance motivation	Other:
objectives of the whole course.	degree	high		middle		low
2 Explain the learning	aspect(s)	Develop knowledge	Improve skill	Boost confidence	Enhance motivation	Other:
objectives of a particular unit.	degree	high		middle		low
3 Explain the	aspect(s)	Develop	Improve	Boost	Enhance	Other:

learning objectives of a particular activity.		knowledge	skill	confidence	motivation	
	degree	high		middle		low
4 Explain the connections between the learning objectives of the whole course, of a particular unit, and of a particular activity.	aspect(s)	Develop knowledge	Improve skill	Boost confidence	Enhance motivation	Other:
	degree	high		middle		low
5 Explain the skills to be grasped in this course.	aspect(s)	Develop knowledge	Improve skill	Boost confidence	Enhance motivation	Other:
	degree	high		middle		low
6 Ask a student to give a news report at the beginning of a class.	aspect(s)	Develop knowledge	Improve skill	Boost confidence	Enhance motivation	Other:
	degree	high		middle		low
7 Use questioning to check how well students have grasped the knowledge.	aspect(s)	Develop knowledge	Improve skill	Boost confidence	Enhance motivation	Other:
	degree	high		middle		low
8 Organize group activities for students to practice public speaking skills.	aspect(s)	Develop knowledge	Improve skill	Boost confidence	Enhance motivation	Other:
	degree	high		middle		low
9 Give guidance while students practice in groups.	aspect(s)	Develop knowledge	Improve skill	Boost confidence	Enhance motivation	Other:
	degree	high		middle		low
10 Ask some students to give a practice speech in front of the class.	aspect(s)	Develop knowledge	Improve skill	Boost confidence	Enhance motivation	Other:
	degree	high		middle		low
11 Ask the class to comment on the speech given by their classmate.	aspect(s)	Develop knowledge	Improve skill	Boost confidence	Enhance motivation	Other:
	degree	high		middle		low
12 Give feedback on one student's speech.	aspect(s)	Develop knowledge	Improve skill	Boost confidence	Enhance motivation	Other:
	degree	high		middle		low
13 Explain the marking criteria for peer assessment.	aspect(s)	Develop knowledge	Improve skill	Boost confidence	Enhance motivation	Other:
	degree	high		middle		low
14 Ask every student to give three required speeches during the semester.	aspect(s)	Develop knowledge	Improve skill	Boost confidence	Enhance motivation	Other:
	degree	high		middle		low

15 Ask part of the class to give a mark for one student's speech.	aspect(s)	Develop knowledge	Improve skill	Boost confidence	Enhance motivation	Other:
	degree	high		middle		low
16 Ask the whole class to give written feedback to one student's speech.。	aspect(s)	Develop knowledge	Improve skill	Boost confidence	Enhance motivation	Other:
	degree	high		middle		low
17 Ask the class to reflect on his/her own performances after giving a speech.	aspect(s)	Develop knowledge	Improve skill	Boost confidence	Enhance motivation	Other:
	degree	high		middle		low
18 Explain to students how they are assessed in this course.	aspect(s)	Develop knowledge	Improve skill	Boost confidence	Enhance motivation	Other:
	degree	high		middle		low
19 Inform students what they are expected to do at the course test(s).	aspect(s)	Develop knowledge	Improve skill	Boost confidence	Enhance motivation	Other:
	degree	high		middle		low
20. Other:	aspect(s)	Develop knowledge	Improve skill	Boost confidence	Enhance motivation	Other:
	degree	high		middle		low

Part V. Overall impression. If you have other comments or thoughts about this oral English course, please write them down here.

This is the end of the questionnaire. Thank you very much!

Appendix 7b. Student End-of-Semester Questionnaire (The Applied Chinese Version)

英语口语学习调查问卷 (学期末)

跟踪编码：_____
第一部分：您的个人信息
性别： □ 男　　□ 女
您是否参加过大规模的英语水平考试（如托福、雅思、四级、六级考试等）：　　□ 是
□ 否
如果参加过，请写出最近一次考试的时间_____　名称 _____分数: _____
所来自的中学： □ 外语学校　　　　□ 非外语学校
开始学英语的年龄： □ 8岁以下　　□ 9-12岁　　□ 13-15岁　　□ 16岁以上
您是否有过出国经历： □ 有　　　□ 没有
如果有，请写出您去的国家: _____　在那里停留的时间: _____

第二部分：在这门课上，您认为在进行如下哪些活动时，是老师在对学生进行评估；学生当时做这些活动的表现会是否影响本门课的成绩。

请根据实际情况在相应的地方画圈。	是老师在对学生进行评估吗？		是否会影响自己本门课的成绩？	
1 是否按时进入教室。	是	否	是	否
2 上课一开始做news report。	是	否	是	否
3 对老师讲解过程中提出的问题进行回答。	是	否	是	否
4 小组活动中发言时，老师在旁边听。	是	否	是	否
5 小组活动后被点名到全班面前讲。	是	否	是	否
6 演讲后对同学和老师的提问进行回答。	是	否	是	否
7 在全班面前做三个规定的打分的演讲。	是	否	是	否
8 对同伴的表现发表看法。	是	否	是	否
9 对自己的在演讲中反映出的优缺点进行反思。	是	否	是	否

第三部分：对自己本学期的进步的感受。

我现在与开学初相比,在如下这些方面	进步的感受				
1 讲英语时的语音语调	进步很大	有些进步	没有改变	有些退步	退步很大
2 讲英语时的流利度	进步很大	有些进步	没有改变	有些退步	退步很大
3 英语口语的词汇量	进步很大	有些进步	没有改变	有些退步	退步很大
4 英语口语的准确性	进步很大	有些进步	没有改变	有些退步	退步很大
5 讲英语时所能表达的内容的复杂度	进步很大	有些进步	没有改变	有些退步	退步很大
6 英语口语交际能力	进步很大	有些进步	没有改变	有些退步	退步很大
7 对用英语进行公众演讲的技巧和策略的了解	进步很大	有些进步	没有改变	有些退步	退步很大
8 明白该如何一步步提高自己英语演讲能力	进步很大	有些进步	没有改变	有些退步	退步很大
9 用英语进行公众演讲的能力	进步很大	有些进步	没有改变	有些退步	退步很大
10 用英语进行辩论的能力	进步很大	有些进步	没有改变	有些退步	退步很大
11 认为自己将来能把英语口语说得很棒的想法	变得更强	有些增强	没有改变	有些减弱	变得更弱
12 认为自己将来能成为优秀的演说家的想法	变得更强	有些增强	没有改变	有些减弱	变得更弱
13 认为自己只要努力就会进步的想法	变得更强	有些增强	没有改变	有些减弱	变得更弱
14 认为自己不论怎么练都不可能有进步的想法	变得更强	有些增强	没有改变	有些减弱	变得更弱
15 用英语进行公众演讲时的自信心	变得更强	有些增强	没有改变	有些减弱	变得更弱
16 用英语进行公众演讲时的焦虑	变得更强	有些增强	没有改变	有些减弱	变得更弱

17 希望自己能在这门课上取得好成绩的想法	变得更强	有些增强	没有改变	有些减弱	变得更弱
18 希望自己是老师眼中优秀的学生的想法	变得更强	有些增强	没有改变	有些减弱	变得更弱
19 担心自己这门课成绩不好的想法	变得更强	有些增强	没有改变	有些减弱	变得更弱
20 努力创造机会练习自己的英语演讲能力	更加努力	有些增强	没有改变	有些下降	非常下降
21 努力扩展自己的英语口语的词汇量	更加努力	有些增强	没有改变	有些下降	非常下降
22 努力改进自己英语口语的准确性	更加努力	有些增强	没有改变	有些下降	非常下降

第四部分：这门公共演讲课上的哪些做法带来了你的进步？

请圈出老师在这门课上使用的课堂做法。	请圈出这些做法对自己有帮助的方面（可多选），并与相应的帮助强度连线起来。					
1 阐明这门课的学习目标。	帮助的方面	增加知识	提高技能	提高自信	提高动机	其它：
	帮助的强度	高		中		低
2 阐明一个单元的学习目标。	帮助的方面	增加知识	提高技能	提高自信	提高动机	其它：
	帮助的强度	高		中		低
3 阐明一个活动的学习目标。	帮助的方面	增加知识	提高技能	提高自信	提高动机	其它：
	帮助的强度	高		中		低
4 阐明不同层次的学习目标之间的关系。	帮助的方面	增加知识	提高技能	提高自信	提高动机	其它：
	帮助的强度	高		中		低
5 阐明学生需要掌握的技能。	帮助的方面	增加知识	提高技能	提高自信	提高动机	其它：
	帮助的强度	高		中		低
6 上课一开始让学生做 news report。	帮助的方面	增加知识	提高技能	提高自信	提高动机	其它：
	帮助的强度	高		中		低
7 通过提问检查学生对相关知识的掌握情况。	帮助的方面	增加知识	提高技能	提高自信	提高动机	其它：
	帮助的强度	高		中		低
8 组织小组活动让学生练习演讲技能。	帮助的方面	增加知识	提高技能	提高自信	提高动机	其它：
	帮助的强度	高		中		低
9 在学生进行小组练习的时候给以指导。	帮助的方面	增加知识	提高技能	提高自信	提高动机	其它：
	帮助的强度	高		中		低
10 请个别学生在全班面前练习	帮助的方面	增加知识	提高技能	提高自信	提高动机	其它：

演讲。	帮助的强度	高		中		低
11 请全班同学对一个学生的演讲进行口头评价。	帮助的方面	增加知识	提高技能	提高自信	提高动机	其它：
	帮助的强度	高		中		低
12 对一个学生的演讲给出自己的反馈。	帮助的方面	增加知识	提高技能	提高自信	提高动机	其它：
	帮助的强度	高		中		低
13 阐明学生在评价同伴的演讲时的评分标准。	帮助的方面	增加知识	提高技能	提高自信	提高动机	其它：
	帮助的强度	高		中		低
14 要求每个学生都做三个规定的打分的演讲。	帮助的方面	增加知识	提高技能	提高自信	提高动机	其它：
	帮助的强度	高		中		低
15 要求部分同学对一个学生的演讲打分。	帮助的方面	增加知识	提高技能	提高自信	提高动机	其它：
	帮助的强度	高		中		低
16 要求全班同学对一个学生的演讲给出书面反馈。	帮助的方面	增加知识	提高技能	提高自信	提高动机	其它：
	帮助的强度	高		中		低
17 要求全班同学对自己的演讲写出书面反思。	帮助的方面	增加知识	提高技能	提高自信	提高动机	其它：
	帮助的强度	高		中		低
18 阐明这门课的考核方式。	帮助的方面	增加知识	提高技能	提高自信	提高动机	其它：
	帮助的强度	高		中		低
19 指明学生在考试中应如何表现。	帮助的方面	增加知识	提高技能	提高自信	提高动机	其它：
	帮助的强度	高		中		低
20. 其它做法：	帮助的方面	增加知识	提高技能	提高自信	提高动机	其它：
	帮助的强度	高		中		低

第五部分： 整体感受。

关于这门课，您如果还有其它方面的感受，请写在下面。

问卷到此结束！非常感谢！

Appendix 8. The Second Journal Topics for Andrew's Class

Thank you very much for having written the first journal. The following are the questions for the second journal, which is about this week's class.

At the beginning of today's class, a girl gave a presentation on how to read people's facial expressions. How did you like her presentation? Did you learn something from her presentation and/or from the teacher's feedback? (今天的课一开始的时候，有一位女同学做了一个关于如何读懂人们的面部表情的报告。你觉得她的报告做的怎么样？你从她的报告中学到了什么？或者从老师的反馈中学到了什么？)

In today's class, the teacher asked you to make up a dialogue and then asked two pairs to perform before the class. How did you like this activity? How difficult was this activity to you? What did you think of the two pairs' performances? Did you learn something from their performances as well as your teacher's feedback? (在今天的课上，老师让你们编一个对话，然后请两组同学到前面表演。你觉得这个课堂活动怎么样？你从同学的表现和老师的反馈中学到了什么?)

Except for the two activities mentioned above, are there any other activities in today's class that have left you a deep impression? Why? (除了上面提及的两个活动，今天课上是否还有其他什么活动给你留下了深刻的印象？为什么？)

Appendix 9. Student Stimulated Retrospective Interview Guide

Instruction to students before the interview:

"The purpose of this interview is to find out how you were engaged in some of the classroom activities in today's class. I will remind you of those activities one by one by giving you a short summary, and after each reminder I will ask you a list of questions relating to what you were thinking, what you were doing, how you were feeling while you were engaged in the activity. If you cannot remember that activity very clearly, I will replay the recording of that part for you."

Generic questions concerning each activity where assessment might be involved:

What did you focus on while you were doing this activity?
Did you enjoy this activity? Why or why not?
How did you feel while you were doing this activity?
How did you like your teacher's comments?
Did you learn something from this activity?

Appendix 10. Student End-of-semester Interview Guide

General feelings about this course

Overall feelings about this course
Any impressive / happy / unhappy moments: describe and explain
Opinions of this course: satisfactory aspects, places for improvement

Achievements and reasons

Achievements: communication skills, knowledge, confidence, motivation, learning strategies, etc.
Reasons: classroom activities, assignments/homework, tests, etc.

Feelings about classroom assessment practices

Feelings about typical classroom activities
Feelings about teacher's feedback in class
Any impressive moments, or any successful / unsuccessful moments

Feelings about the final test:

Preparation, experiences of the test, if induced anxiety, if affected self-confidence, perceived learning after the test, opinion on the format, content, and marking, etc.

Feelings about any other assessments that contributed to the final score:

When, format, weighting, if induced anxiety, etc.

Appendix 11. Transcription Notation

T	Teacher
A1	An identified student
S1	An unidentified student
SS	Several or all students simultaneously
[…]	Omission of elements not necessary for the current analysis
(())	Researcher comments including those concerning non-verbal actions [e.g. ((laugh))] or translation of Chinese words [e.g., 一个预告片((yi ge yu gao pian, a trailer))]
<xxx>	Unintelligible speech
...	Long pause
_	Speaker emphasis (e.g. It is a very good job.)
-	Truncated word, or unfinished sentence (e.g. I- I want to)
?	Rising vocal pitch or intonational contour
::	Lengthening of a word or sound (e.g. Can I look at your::draft?)
' '	Speaker quoting other people's words

Appendix 12. Andrew's Course Description

College English Speaking Level 3 (S3)

For Non-English foreign language majors

1. Objectives

This is a level-3 course of **English Speaking** for non-English foreign language majors.

The objectives of the course are

to encourage and enhance students' self-confidence in expressing themselves in English;
to improve students' fluency and accuracy in speaking English;
to help students to enlarge their English vocabulary and knowledge of the English cultures;
to help students to learn to talk about various topics related to daily life and social issues;
to provide opportunities to practise skills needed in conversations, story-telling, presentations, etc. in English.

2.   Main Text

New StandardCollege English (Level 2)*: Listening and Speaking* (**NSCE-LS2**, for short), chief editors: Simon GREENALL &WEN Qiufang. Teacher's Book, prepared by JINLixian, Martin CORTAZZI &PhilipLEETCH. Beijing: Foreign Language Teaching and Research Press. 2009.

3.   Hours & Credits   2 hours per week; 2 credits.

4. Schedule

Week 1– 2 Sept. 6th ~ 17th	Course Introduction **Unit 1 College Life** (*NSCE-LS*2:U1 College Culture)
Week 3 – 4 Sept. 19th ~ Sept. 30th	**Unit 2 Feelings**(*NSCE-LS*2:U2 Mixed Feelings) Mid-Autumn Festival Holidays in Week 3: three consecutive holidays Wed + Thurs. / Fri. Please adjust your teaching plans accordingly: September 19th (Sat.) is a working day on the Thursday teaching schedule. September 25 (Sat.) is a working day on the Friday teaching schedule.
Week 5 Oct. 1st – Oct. 10th	National Day Holidays
Week 6 – 7 Oct. 11th ~ Oct. 22nd	**Unit 3 Crime Watch** (*NSCE-LS*2:U3 Crime Watch)
Week 8 – 9	**Unit 4 News**(*NSCE-LS*2:U4 News)

Oct. 25th ~ Nov. 5th	
Week 10- 11 Nov. 8th ~ Nov. 19th	Unit 5 Leisure and Hobbies (*NSCE-LS*2:U6 Sporting Life)
Week 12 – 13 Nov. 22nd ~ Dec. 3rd	Unit 6 Nature and Animals (*NSCE-LS*2:U7 Animal Planet)
Week 14 - 15 Dec. 6th ~ Dec. 17th	**Unit 7 Jobs and Career**(*NSCE-LS*2:U9 Have you Got What it Takes?)
Week 16 Dec. 20th ~ Jan. 24th	Final Exam

5. Course Assessment

Attendance: 10%
Participation: 30%
Presentation: 10%
Exam: 50%.

6. General guidelines for group presentations by students (10%)

A student presentation may be conducted as an individual, pair, or group presentation.
Each presentation must have a central theme, relevant to the course.
Each student must speak up during the presentation.
All presentations should (preferably) be scripted and must be timed (2-3 minutes for each student).
The presentation is a practice of public speaking and teamwork, and will account for 10% of the course assessment. Group presentations are encouraged in big classes.

7. Exam (50%)

The exam is given as oral tests, in which students will talk to each other in pairs on ONE given topic. The topic is to be picked by lucky-draw before the exam; then the student will have 5 minutes to prepare. Each oral exam session will last 3-5 minutes. All the topics are related to the course contents.

Question sample: *What type of student do you think you are?* (Ref. Unit 1 Talking point, page 5)

Conversation sample: * Below is a sample for groups of 2-3.

Topic No. 1 You are <u>A</u>.		Topic No. 1 You are <u>B</u>.		Topic No. 1 You are<u>C</u>.
The ***Campus Theatre Night*** will be on New Year's Eve! You and your classmates are putting on a play (say, *Mama Mia*) and need to book a room for dress rehearsal. Discuss the plan with your partner. <u>You offer to seek support of the Students' Union and</u>		The ***Campus Theatre Night*** will be on New Year's Eve! You and your classmates are putting on a play (say, *Mama Mia*) and need to book a room for dress rehearsal. Discuss the plan with your partner. <u>You offer to contact the staff of the uni auditorium</u>		The ***Campus Theatre Night*** will be on New Year's Eve! You and your classmates are putting on a play (say, *Mama Mia*) and need to book a room for dress rehearsal. Discuss the plan with your partner. <u>You offer to invite a teacher to act as</u>

secure a room with multi-media facilities. **Agree on your plan.**	to secure a time slot to rehearse on stage. **Agree on your plan.**	your language advisor. **Agree on your plan.**

8. Criteria(for both presentations and exam)

Students' scores will be awarded according to their performance in the following three aspects:
Pronunciation and Language Quality (40%): (Whether the students speak fluently and with correct pronunciation and grammar; whether they possess an appropriate and effective variety of vocabulary and can express their ideas in flexible sentence structures.)
Organization and Content (40%): (Whether the ideas are conveyed in an organized way and cover the necessary aspects of the given topic.)
Performance and Teamwork (20%): (Whether the students speak with proper manners and cooperate with their partners well.)
The total score is 100 %.

Appendix 13. Introduction to the 9 Sections of Each Unit in Andrew's Textbook

Real Communication: Listening and Speaking

Starting point is a pair- or group-work activity which introduces the unit theme.

Inside view provides listening practice by means of a video story of three students, Mark (English), Kate (American) and Janet (Chinese) at the University of Oxford in England. Filmed on location in Oxford, it shows their typical lives, interests and concerns, and provides an insight into the university and the city. There are comprehension activities and a **Language and culture** box to explain unfamiliar cultural references. **Everyday English** highlights some of the words and expressions which are very common in spoken, contemporary English, and which may be difficult to understand. The section ends with a guided functional dialogue, with a box of useful functional expressions, taken from the video story as references.

Talking point is a pair- or group-work activity which remains close to the unit theme, but allows a change of pace in the lesson.

Outside view uses short extracts from video material in which the English is authentic and roughly graded to the students' level. The accompanying activities are designed more to enable the students to understand the main ideas, rather than to check detailed comprehension. There is a section for **Developing critical thinking**, with a similar intention to the ones in the Integrated Course.

Listening in contains two listening passages which provide further practice listening to roughly graded material. As with Outside view, the intention is to expose the students to language which may be slightly higher than their present level of English, but which will prepare them for listening and understanding in real-life contexts. This section finishes with a **Developing critical thinking** activity too.

Presentation skills includes advice on techniques for giving effective presentations, as well as a box of functional expressions. The main activities lead the students to give a presentation related to the unit theme.

Pronunciation focuses on the specific points in pronunciation, stress and intonation which cause Chinese speakers of English difficulty, and includes aspects such as linking sounds, stressed words, and sense groups.

Unit task contains a task which allows the students in pairs or groups to review all the language skills they have covered during the unit.

Unit file is a summary of the language points and skills presented in the unit.

Appendix 14: The Minutes of the Teacher Group Meeting for the Final Exam (Andrew's Case)

Minutes in Chinese	English translation
昨天集体备课关于期末考题的意见汇总如下：	The following is a summary of our yesterday's discussion on the final exam.
考试各班随堂，允许学生提前结伴，允许学生提前了解考题，考试时抽签决定考哪一道题。	The final exam will be held during regular class time. Students are allowed to form pairs before the exam. Students will be informed of the exam topics in advance. Students will draw lots to decide on the topic at the exam time.
题型用对话，两人一组（单数时允许三人组）。	The exam will take the form of a situational dialogue, with two students taking the exam together. If there is an odd student in a class, this student can take the exam with another pair.
出题重点参照课本Inside View中的话题和会话功能。	The design of the exam topics should be based on the topics and the language functions covered in the Inside View section of each unit in the textbook.
每单元出一道题，每套题三个角色卡（以便三人组使用）。	We should design one exam topic for each unit and three roles for each topic just in case three students have to take the exam together.
大家分工出题，第一单元就用大纲里的那一道，其他单元的题，请尚未担任主备的老师们辛苦一下。	We are going to divide the work of designing the exam topics among us. We will use the one listed in the Course Description as the topic for the first unit. Those teachers who have not prepared the lesson plan for one unit during the semester will design the exam topic for one unit.

Appendix 15: Andrew's Final Exam

College English Speaking Level 3 Final Exam

Instruction for the students:

The test will be conducted during regular class time in the week of December 21-25.
All topics are related to the course.
You will talk to each other in PAIRS on ONE of the given topics, picked by lucky-draw about 5 minutes before you take your exam.
Each pair should have a conversation for 5-6 minutes.
Each student will be marked on the basis of the individual performance, in accordance to the "Marking Criteria".
Warning: Failure to take part in the exam at the appointed time will result in a failing mark on this course. In case of an emergency or sickness, students who miss the exam may apply for a make-up session at the Main Office of the SESP.

Procedure:

Preparation Week (December 14-18):

Students will see all the topics, sign up as pairs on the "Name list and Scoring Sheet".
Allow groups of three in case of odd numbers of students in the class.

Exam Week December 21-25.

Students should arrive in time to prepare and to take the exam.
During the exam, students should have a conversation on their own.
The teacher will not be involved in the conversation.

Marking Criteria (for both presentations and exam):

Students' scores will be awarded according to their performance in the following three aspects:
Pronunciation and Language Quality (40%): (Whether the students speak fluently and with correct pronunciation and grammar; whether they possess an appropriate and effective variety of vocabulary and can express their ideas in flexible sentence structures.)
Organization and Content (40%): (Whether the ideas are conveyed in an organized way and cover the necessary aspects of the given topic.)
Performance and Teamwork (20%): (Whether the students speak with proper manners and cooperate with their partners well.)
The total score is 100 %.

Topics[10]:

TOPIC 5   What do you do for sport?

Topic No. 5 You are A.		Topic No. 5 You are B.		Topic No. 5 You are C.
An American sports delegation will come to visit our uni. **You are one of the Student Union leaders** who are assigned to accompany the delegation members. You need to: Describe to them the sports you can play at your college, including non-competitive sports, such as swimming and badminton, etc.		An American sports delegation will come to visit a Chinese university.**You are one of the delegates.** You need to: Listen attentively to the introductory comments given by your Chinese hosts on sports at uni and give polite and appropriate responses. Ask some questions on the types of sports activities on		An American sports delegation will come to visit our uni. **You are one of the Student Union leaders** who are assigned to accompany the delegation members. You need to: Talk about what you know about the most popular sports in the US, such as football, baseball, or basketball and NBA games.

[10] There were altogether seven topics, one for each of the seven units covered in the semester. Here to save space, only topic 5 and topic 7 were included, which were related to the lessons that were observed in this study.

Describe the sports you like and how you can join a team on campus.	campus you are visiting.	Say you like Yao Ming very much and give the reasons.

Topic No. 7   What do you want to do after finishing college?

Topic No. 7 You are A.	Topic No. 7 You are B.	Topic No. 7 You are C.
Choose ONE of the professions below as your future career and tell your friend(s) about it. An office Administrator An education manager A travel guide	Choose ONE of the professions below as your future career and tell your friend(s) about it. A simultaneous interpreter A lawyer A free-lance writer	Choose ONE of the professions below as your future career and tell your friend(s) about it. A business manager A public relations officer Music critic

Appendix 16: Unit 6 *Sporting Life* in Andrew's Textbook

Starting point

1 Work in pairs. Read the suggestions and decide which sports they refer to.

Many people think some sports, like football and tennis, have become too predictable or boring. Here are some suggestions about how some sports could be improved.

1 "I would make them use a circular pool, and make them try to pass each other."

2 "I would let the referee look at a video before he decides to send a player off."

3 "I wouldn't let people over two metres tall play, or they could put the basket higher up."

4 "Why don't they make them fight in a round ring, not a square one, because at the moment it's too easy to get trapped in the corner?"

5 "I'd let them all go down the slope together, racing against each other, and not against the clock."

6 "If the goal posts were wider apart, players could score more often."

7 "Make them use heavier rackets. This would mean more accuracy and less power."

2 Work in pairs and discuss the suggestions. Which ones would improve the sports and which wouldn't?

Inside view

Conversation 1

1 Look at the photos and answer the questions.

1 What's happening?
2 Where does it take place?
3 What do you think the rules are for this kind of race?
4 Where are Mark, Kate and Janet?
5 What has Mark just done?
6 How do you think they feel?

2 Watch Conversation 1 and follow the map.

Language and culture

Every Oxford college has a **rowing** team, and there are rowing races between the colleges every term.

Each college has a **boathouse** on the river, where the rowing boats are kept. There's also a bar and somewhere to watch the racing.

The main **river** which runs through Oxford is the **River Thames**, which then flows to London and the North Sea.

62

305
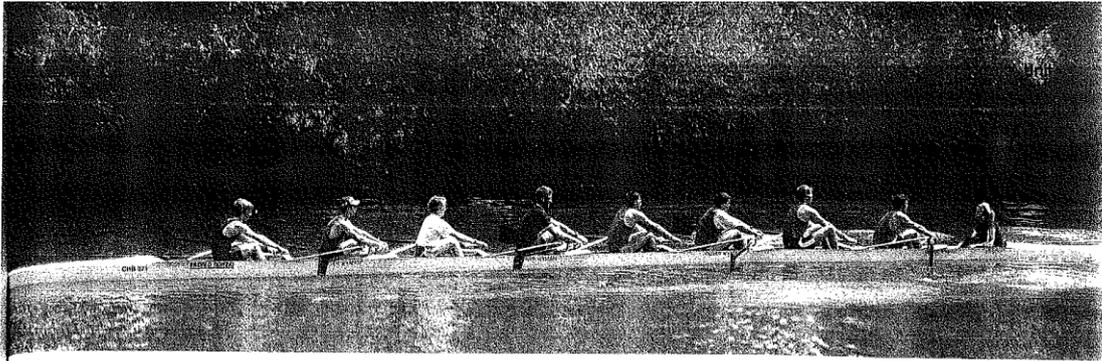
**Conversation 1**

Catte /kæt/ Street 卡替街（牛津大学哈福特学院所在街道）

St Aldates /ˈɔːldeɪts/ 圣阿尔代街（牛津大学基督学院所在街道）

Christ Church College 牛津大学基督学院

Folly /ˈfɒli/ Bridge 福利桥（位于牛津市中心南边，北接圣阿尔代街，横跨泰晤士河）

Conversation 2

bump /bʌmp/ v. （使）撞击，（使）猛击
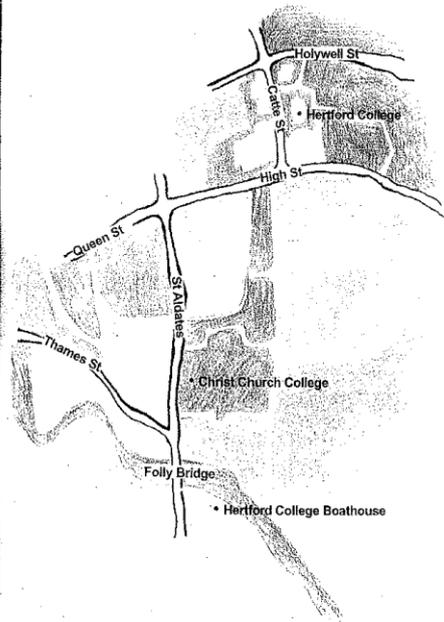
3 Watch Conversation 1 again and choose the best way to complete the sentences.

1 The conversation takes place _____.
(a) in college before the racing begins
(b) on the river bank during the rowing races
(c) in the boathouse after the practice race
(d) when Kate and Janet walking along the river on the way to the boathouse

2 If Mark rows well in the practice race _____.
(a) Janet will come and watch later
(b) he'll get a place on the college team
(c) the college team will row well today
(d) Kate will go and watch

3 Kate's plans for the afternoon include _____.
(a) having lunch at the boathouse and watching the rowing
(b) doing an essay and going to the river
(c) showing Janet where the college boathouse is
(d) taking Janet to the boathouse and having lunch

4 When Kate tells Janet her plans to go to the boathouse, Janet _____.
(a) knows where it is
(b) doesn't know how to get there
(c) says she doesn't want to come
(d) knows her way to the river

Conversation 2

4 Look at the map and decide where the boat race takes place.

5 Watch Conversation 2 and answer the questions.

1 What does Mark's boat manage to do?
2 What does Mark hope to do?
3 How seriously did Mark hurt himself as he was getting into the boat?
4 How does Mark compare with other people who rowed with him?
5 How do people know if they have got a place on the team?
6 Why is Kate pleased that Mark has got a place on the team?

·63·

306

6 Watch Conversation 2 again and complete the sentences.

Kate   So the rules are ... the boats follow each other and the one behind has to bump the one in front ... just like that one has done.

Janet Is that Mark's boat?

Kate   Yes! Look, his boat is about to bump the one in front! (1) _____!

...

Mark   Hi you guys!

Kate   Fantastic, Mark. (2) _____ _____!

Mark   Well, we won the practice race, but (3) _____ _____ getting a place on the team. (4) _____ _____ there are at least three other people on the team who have rowed before. (5) _____ _____ they were better than me.

Janet Don't worry, Mark. (6) _____ _____.

Mark   And then I hurt my knee getting into the boat.

Janet (7) _____!

Kate   Too bad, but it's only a scratch. Listen up, Janet is right. (8) _____ _____, Mark. You were the strongest looking guy in the boat today. Chill out!

Mark   Hey, they're putting the team list on the door.

Janet Let's go over and see.

Mark   No, you go! I can't bear to look!

Kate   OK.

...

Kate   Hey, Mark, great news! You got a place on the college team!

Janet (9) _____!

Kate   That's great, Mark, (10) _____. You trained so hard.

Mark   I can't believe it!

Everyday English

on your way
Yes, got it!
No problem.
Listen up!
Chill out!
I can't bear to look!

7 Work in pairs and answer the questions about Everyday English.

1 **on your way** Does this mean (a) going right now, or (b) planning to go?

2 **Yes, got it!** Does this mean Janet (a) understands, or (b) doesn't understand?

3 **No problem.** Is this (a) an expression to reassure someone, or (b) simply a response to thank you?

4 **Listen up!** Does this mean (a) listen to me, or (b) be quiet?

5 **Chill out!** Does this mean Kate wants Mark to feel (a) excited, or (b) relaxed?

6 **I can't bear to look!** Does this mean Mark (a) wants to look, or (b) doesn't want to look?

8 Work in pairs and act out the conversation.

Student A Ask where Student B is going.

Student B Say that you're going to a sports event and invite Student A.

Student A Explain that you're too busy but you'd like to come later. Ask where Student B will be.

Student B Sympathize with Student A and give directions. Explain that you / your team won the last match.

Student A Congratulate Student B.

Student B Express concern that you / your team may not win the next match.

Student A Reassure Student B.

64·

Sympathizing

That's too bad!

I know how you feel.

I'm so sorry!

Giving directions

Go down ...

Turn right / left into ...

Walk / Go along ...

Cross·over ...

Keep going ...

... are on the right.

It's the last one along.

You can't miss it.

Congratulating

Well done!

Fantastic!

You were amazing!

Great news!

Congratulations!

That's great!

You deserve it.

Expressing concern

I'm worried about ...

The problem is ...

I can't help thinking ...

Reassuring

Don't worry.

Everything will be OK.

No need to get nervous.

Chill out!

Talking point

1 Work in pairs.

2 Work together and complete the table.

	Your partner	You
favourite sports		
favourite sports team		
international players		
highest scoring player		
college sports		
favourite spectator sports		
greatest college sports teams		
best match		
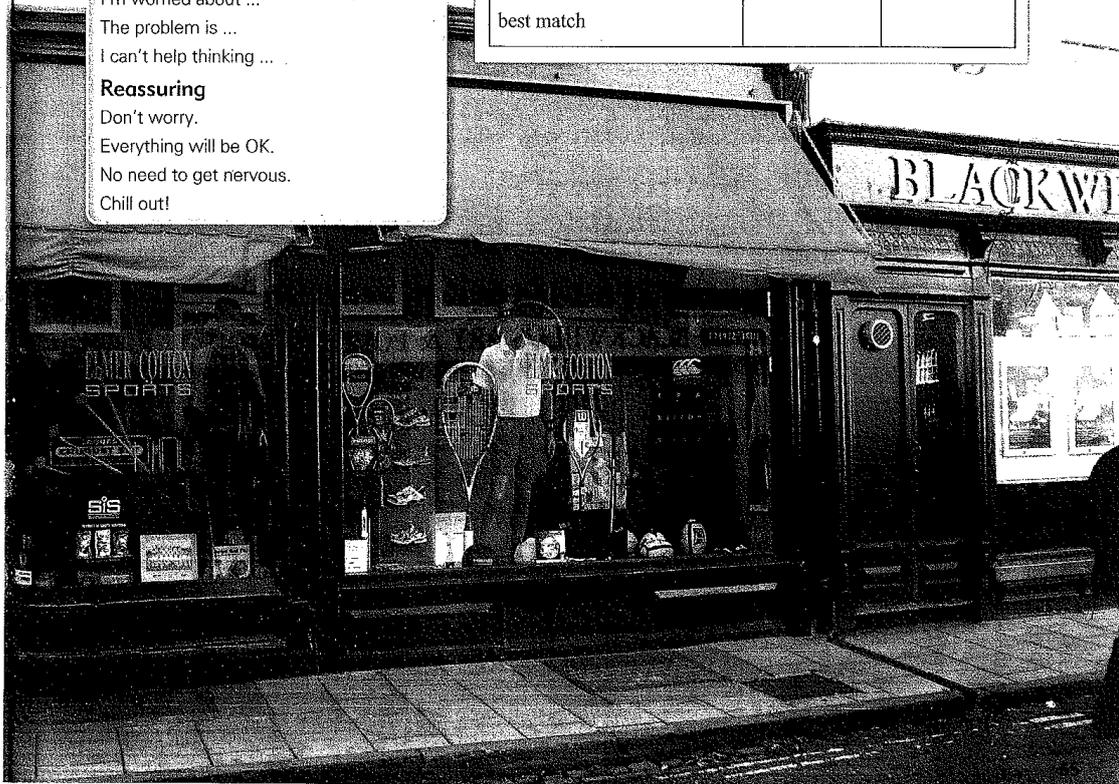
Outside view

1 Look at the photos. What do they tell you about the rules and features of Tae Kwon Do?

Tae Kwon Do /ˈteɪˈkwɒnˈdəʊ/ *n.* 跆拳道
originate /əˈrɪdʒəˌneɪt/ *v.* 发源，开始
formulate /ˈfɔːmjʊˌleɪt/ *v.* 构想、制定（计划、制度、建议）
martial art /ˌmɑːʃl ˈɑːt/ *n.* 武术（如空手道、柔道或功夫）
punch /pʌntʃ/ *v.* （用重拳）猛击，揍
flexibility /ˌfleksəˈbɪləti/ *n.* 柔韧性，灵活性
Dan /dæn/ *n.* 段（柔道、空手道、围棋等运动员的等级）
smack /smæk/ *v.* （用手掌或扁平物）掴，打，拍
pad /pæd/ *n.* 垫，衬垫
release /rɪˈliːs/ *v.* 释放，放出
alternatively /ɔːlˈtɜːnətɪvli/ *ad.* 或，或者，要不
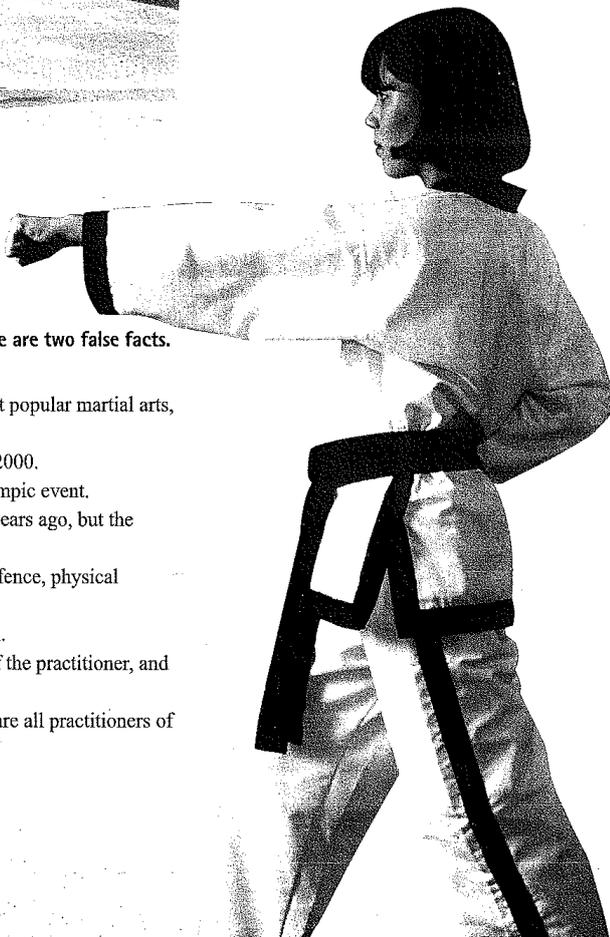
2 Read some facts about Tae Kwon Do. There are two false facts. Which are they?

1 Tae Kwon Do is one of the world's most popular martial arts, with practitioners all over the world.
2 It became an official Olympic event in 2000.
3 It's the only martial art which is an Olympic event.
4 Tae Kwon Do originated thousands of years ago, but the modern rules were formulated in 1955.
5 It combines fighting techniques, self-defence, physical exercise, meditation and philosophy.
6 Using your hands and arms is forbidden.
7 The colour of the belt shows the rank of the practitioner, and the highest rank is black.
8 Men, women and even young children are all practitioners of Tae Kwon Do.

66·

309

Watching and understanding

3 Watch the video clip and complete the answers to the interviewer's questions.

1 I wonder if first off you could explain a bit of the history of Tae Kwon Do.
  Tae Kwon Do originated about
  _____ in Korea.

2 How does it differ to other martial arts?
  Tae Kwon Do is basically the _____
  martial art.

3 What kind of mixture of people do you get coming along?
  It's open to _____, we take any flexibility level, any age, any _____.

4 How long have you been in this sport?
  _____, and I'm a second Dan instructor now.

5 What got you into it in the first place?
  I've always wanted to do martial art, I lacked a little bit of _____ I must admit.

6 What do you think people in general get out of it?
  It's _____, as well as _____.

7 Whereabouts do you hold your classes?
  _____ are Long Crendon here on a Tuesday and Thursday.

8 How would they go about finding out?
  The easiest way of finding out is _____
  which is www.bytomictaekwondo.com.

4 Watch the video clip again and answer the questions.

1 Who created Tae Kwon Do in recent years?
2 Who was it created for?
3 Who else in the instructor's family had done Tae Kwon Do?
4 How does Tae Kwon Do release the stresses of everyday life?
5 What does the instructor offer the interviewer?

> ### Listening to natural English: native speaker grammar
>
> In spoken English, it's quite common for speakers to make grammatical mistakes or for sentences to be unfinished.
>
> *It's in the present form we train and teach now is formulated in 1955 by a Korean general.*
>
> *Well hopefully, it's fun, as well as learning something.*
>
> *All the information is on the website or speak to the office.*
>
> It's rare that these mistakes create comprehension difficulties for the listener. So if we make some mistakes as we speak, there's no need to worry – remember that native speakers of English make mistakes too.

Developing critical thinking

5 Work in pairs and discuss the questions.

1 What are the advantages of learning a martial art?
2 Do you think everyone should learn self-defence?
3 At what point does self-defence become aggressive behaviour towards others?
4 What other sports are fun and help make new friends, as well as develop fitness, flexibility and beat stress?

·67

310
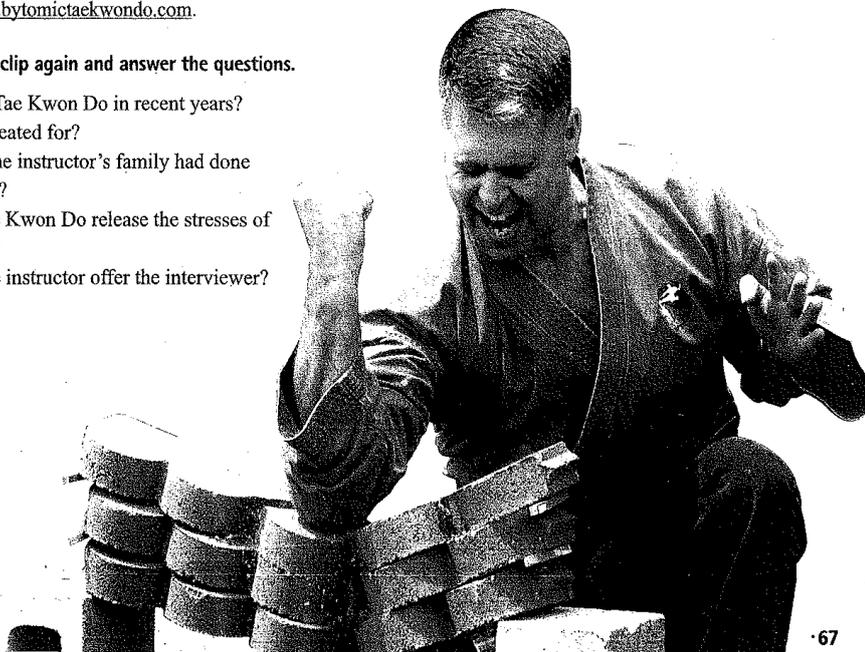
Presentation skills

Holding an informal discussion

1 Work in pairs and look at these two opinions:

Great sportsmen and sportswomen should be paid as much as film stars.
Sport is for amateurs. No one should expect to be paid.

Which opinion do you agree with?

2 Talk about your opinions about professional sport. Think about:

* NBA players
* Olympic athletes
* Premier League football players
* professional tennis players
* Formula One drivers

Which sportspeople deserve to earn large sums of money? Make a note of your opinions.

3 Work in groups of three or four and hold a discussion.

4 Work with the whole class, and continue the discussion. Find out what the most common opinion about payments for sportspeople is.

Holding an informal discussion

Usually, an informal discussion is not prepared in advance. It's a natural part of a conversation.

Take turns to speak and try to make sure everyone has a chance to give their opinion.

We can show that we want to say something by using our body language, eye contact, sounds such as *Er ..., Um ...*

We can express strong opinions, but try to be light-hearted.

If we're sure about something factual or a logical conclusion, we can say so.

We should try to pick up from what someone has already said. It doesn't have to be the very last thing, but it has to be relevant to the overall theme of the discussion.

We can interrupt, but we should try to do it politely, during a natural pause. We should also try not to speak while other people are speaking.

Expressing strong opinions
I really think ...
I certainly don't think that ...
It's nonsense to say ...

Expressing certainty
Actually / In fact, ...
Clearly / Obviously, ...
There's no doubt ...
Surely, ...

Inviting others to speak
What do you think, Janet?
What's your take on all this?

Entering the discussion
That's a good point. I just want to add that ...
Well, I take your point, but ...

70

311

3 Watch the video clip and complete the answers to the interviewer's questions.

1 I wonder if first off you could explain a bit of the history of Tae Kwon Do.
Tae Kwon Do originated about
_____ in Korea.

2 How does it differ to other martial arts?
Tae Kwon Do is basically the _____
martial art.

3 What kind of mixture of people do you get coming along?
It's open to _____, we take any
flexibility level, any age, any _____.

4 How long have you been in this sport?
_____, and I'm a second Dan
instructor now.

5 What got you into it in the first place?
I've always wanted to do martial art, I lacked a
little bit of _____ I must admit.

6 What do you think people in general get out of it?
It's _____, as well as _____.

7 Whereabouts do you hold your classes?
_____ are Long
Crendon here on a Tuesday and Thursday.

8 How would they go about finding out?
The easiest way of finding out is _____
which is www.bytomictaekwondo.com.

Appendix 17: Unit task of Unit 9 in Andrew's textbook

UNIT TASK
Interviewing candidates for a part-time job / an internship

1 Work in groups of four.

Pair A You are going to interview two candidates (Pair B) for a job. Decide on:
- the type of job
- the qualifications, experience and skills needed

Pair B Listen to Pair A for their discussion. Then work on your own and decide what you will say in the interview.

Pair A Discuss the questions you will ask in the interview.

2 Do the following:

Pair A
- Interview the two candidates in turn.
- Decide who should have the job.
- Tell Pair B who got the job, giving reasons for your decision.

Pair B
- Discuss how you thought you did in the interview.
- Listen to Pair A. Say how you feel and ask questions.

Appendix 18: Exercise 7 of Unit 9 in Andrew's Textbook

7 Work in pairs and act out the conversation.

Student A	Ask Student B if they know what they want to do after finishing college.
Student B	Say you're not sure and talk about different possibilities.
Student A	Listen and comment.
Student B	Say more about your ideas.
Student A	Check that something you know about Student B is correct.
Student B	Confirm or correct what Student A has said. Ask about Student A's career plans.
Student A	Say that you think it's important to plan ahead and that you know what you want to do. Talk about your plans.
Student B	Listen and comment.
Student A	Say that there's a careers fair on and suggest going together as it might help give Student B some ideas.
Student B	Agree to go.

Appendix 19: Classroom Recording Transcripts Concerning Mary's Explanation about the Final Test

M: So, I want to give you a little lecture on Christmas, since it is Christmas season. But before we do that, before we go on any further, I want to give you the schedule for the exam. It will be next week, ok, at the same time. I want you to sign up, you should sign up with a partner. So two of you should come together. So just put your name on each line, ok, and yeah so then you don't have to come for the whole class. You just come at the time when you are scheduled, yeah, or you know, maybe five minutes beforehand, in case I am ahead of time. But never am I ahead of time in my final exams, so I probably won't be this time either. You should be a little bit late, you know. So each pair will have 10 minutes together, and we are going to, let me pass it ((the sign-up sheet)) around while I am talking. … So you are just top part ((looking at the student who had just put down their names in the first time slot)), yeah, 8am Tuesday. You need a partner, so decide who you are going to come in with. … And ok, so the exam, the final exam will be, I have decided to narrow it down, rather than try to talk about all the things we've talked about this semester. I'm gonna have two topics: one is going to be the film we just watched, and I want you to talk about what the film meant to you, what kind of lessons there were in it for you, uh what moved you, what didn't move you, what you like, what you didn't like, whatever any reaction to the film. The second thing will be the speeches you guys impersonated, do you know? So you guys , you know, you gave someone else's a speech, right? That's the one I am talking about, not your own speech.

That's the second topic. So I could ask you to talk about the speech, what you got out of the topic of the speech, ok, so what moved you, what was the meaningful message in the speech. So I really wanted to talk about those speeches in class, but we, because we were trying to get through with everyone's speech, we never had a chance to talk about it, and so uh, this is the opportunity, to talk about the meaning of the speeches. Sorry you guys ((two students came into the classroom)) are kind of late. You missed out very important information. So anyway, do you understand that? So I might say 'ok, tell me about your speech and who gave it. And you know you don't have to give the speech again, and I don't want you to memorize anything. If I think you've just coming in and spoken something you've memorized, you'll get a low score. I don't want something memorized. Do you understand what I'm talking about? I want you to talk to me or talk to your partner, ok, so maybe we'll have a three-way conversation. You understand? So we just gonna talk either talk about the movie or we're gonna talk about the topic of your speech. Do you understand that? Does that make sense?

S1: So you are going to choose?

M: I'm gonna choose. I'll probably do the choosing. So you do have to prepare two things, prepare to talk about either one of those things. So I will let you know when you come, yeah, which one you should talk about.

S2: Should I see the film again because I didn't see the first part?

M: Yeah, probably you should, yeah, if you want to prepare for it, yeah, you'd better. yeah. Ok, so any questions? Does that make sense?

S3: Well, if you choose the second topic, and we have prepared for different speeches?

M: Yeah, because there are two of you, I would ask each of you to talk about a short bit-

S3: -talk about the piece of speech XXX?

M: No, the piece of speech you gave, yeah, because I don't think you can remember everyone else's speech. What's that? ((indicating another student who wished to talk))

S4: Any speech we like?

S5: No, the speech you gave.

M: Yeah, the speech you gave, so yeah, because you are familiar with it. I think you should be able to talk about that easily, because you have remembered it. You are familiar with it. So I don't want you to re-give the speech. But I do want you to tell me what kind of impression the message of the speech had on you, or what kind of impression you think it had on the people, what kind of speaker that person is, you know just your reaction, your general reaction to that speech.Was it moving? How did it move you? That kind of thing, ok. So anybody do political speeches. How many of you did political speeches?

S2: Obama.

M: Obama, ok, yeah, ok, now if you find that a problem, you know, because it's about, you know some of the topics are political and they may not, or you may not quite understand them, or something, I don't know. But anyway, just I would like you to talk about what moved you about your speech. Ok.

S6: Do we have to talk about the speeches we imitated or the one we gave?

M: The speech you imitated.

S6: And the movie must be *Christmas Carol*?

M: Yes. yes. So we just talked about it today. We just talked about, I think we've pretty much covered a lot of the themes in the movie. So I want you to talk about what it meant to you. what were the lessons you think that it was teaching, that kind of thing, yeah. And some of the comments that were made today, those are, all those comments were good, so yeah, again, I do not want something memorized. So I don't want you to go online and find somebody else's explanation about the movie and memorize it and tell me what they said. I want you to talk to me, ok. So, if it's not real communication, you'll get a lower score. Understand?

S1: So will it be very formal? Like an exam? or we can just have a chat?

M: We'll have a chat. Yeah, it will be a chat. Yeah, it's a little stealthy because you are coming in and I am grading you. So that's always nerve-racking, but no, just be a conversation, yeah, that's, we'll have a conversation.

S2: The three of us?

S7: Between the three?

M: That's right, yeah, there is three of us. yeah.

S7: Would you like to ask some questions or we just organize the talk?

M: Well, I ask questions. I could, as we go, I guess the question would be 'what did you think of your speech', 'give me your reaction to your speech', and then you'll talk. So I don't need to talk during the test, I mean I may ask questions to direct you, but I need to see, I need to pay attention while you are talking, because I need somehow evaluate you, although I've evaluated you pretty well already. I think, but so I won't do much talking.

S8: I don't think we have to recite something?

M: No, you don't. I don't want you to recite something.

S8: Just review maybe the topics?

M: Yeah. so if you, yeah, just think about it, think about it ahead of time and be prepared to come and talk about it, that's all.

S8: Maybe we have to find some background of the movie or the speech? Do we have to do that?

M: No.

S8: So we just listen to the original version of the speech and try to understand that?

M: So you've already did that, right? You've already listened, I mean you've already done your speech. So I assume you are pretty familiar with it. Yeah, so I just want you to talk about what the content of the speech meant to you. And I think even with the political speeches, you can still do that, you know, even if you don't understand the particular things, maybe you can't relate to the particular things they were talking about, I think you can stil give a reaction to the speech, to the man, whatever.

S8: Why we chose this to imitate?

M: Yeah. Ok.

S2: Can I say something about his technique in the speech?

M: Sure. You can talk about anything you want regarding the speech, but yeah, say something about the content, I like you to mention something about the content, ok. So I , this is going to give the chance to reuse a lot of vocabulary perhaps you learned in that speech, right? Yeah, so and just to give me the sense that you can talk about that issue. So I think it's a fairly fair you know exam, because something you've covered, and so I am just seeing if you can talk about something that you've already that we've covered in class, right? Yeah. Does that make sense? Ok, any other questions? … No. …Ok, let's talk about, I just want to give a little introduction to Christmas.

Appendix 20. Mary's Grading Process at the Final Test

The first pair talked about the film, and during their test, Mary asked them some very hard questions. After their test, Mary said "Oh, grading is tough. It's a standard in this university: you'll have to have only 40% above 85; for the rest , also a lot of 84s. But anyway, I would guess Cathy I would give her 80. she did well but she had a lot of problems, a few problems with the pronunciation and stuff. Andrew I think did very well, yeah, though a little bit shy. Andrew, he is very smart, a little shy and a little quiet, but that's not a problem. So I probably give him an over 85. Yes, I am just not sure. I need to compare with the other students". Later she added that "I thought Andrew's content was much more in-depth than Cathy's". The second pair of students talked about their speeches. After their test, Mary said, "Well, Vivian, Vivian's English of course is very good. Yeah, she's been in America for a while, so she got a high score. Of course Lily, she is a little choppy but she was able to keep going, and I thought that was really good. Yeah, she wasn't afraid to just continue, but obviously she has some grammar problems and pronunciation problems, I couldn't quite understand her, so she got below 80, or 78 to 80". She also added that "I noticed, Lily, she talked a lot, but the depth wasn't that great.…Vivian, she could use some higher vocabulary, but she just speaks very natural informal English".

Appendix 21. Linda's Course Description

Public Speaking Syllabus

FOR STUDENTS: 1st year 2nd semester (March – July 2011)
CLASSROOM HOURS: 3 hrs/week (Classes 1-8);

Course Description

Good speaking is good thinking — reasonable, well-informed, creative, and flexible. This course is designed to cultivate the students' ability to speak effectively in public, with a clear sense of purpose, resourceful thinking, and confidence to express ideas. Special consideration is given to listening behaviors and the ethical conduct of speech in various occasions. After completing this course, the students should be able to deliver various types of speeches, critically evaluate speeches of others, and approach public speaking with greater confidence.

Required Texts

The Art of Public Speaking (**PS**) **10th edition**, by Stephen E. Lucas, Beijing: Foreign Language Teaching and Research Press, 2010. (inclusive of one disc)
Oral English (**OE**) *2, Contemporary College English,* by Yang Limin, Beijing: Foreign Language Teaching and Research Press, 2005.

Course Organization

The 15 weeks for the course are classified as **12** practice weeks and **3** speech weeks.

For a **practice week**, the instructor will lecture on the skills of public speaking following the course schedule. Alternatively, the instructor may give the students a test on the assigned PS chapters for required reading. The rest of the time the instructor can show sample speeches and ask students to give critical comments. Then the students will have time to practice the focal speaking skills of the week in class and get feedback from their peers and the instructor. The OE book will serve as a resource book, which provides ideas for speeches and language input.

For the three **speech weeks**, each student will prepare and deliver three speeches. All class hours of the week will be devoted to students' speeches and on-site peer evaluation, plus listeners' oral feedback in three-hour-a-week classes. All students must turn in their preparation outlines (a hard copy) for a given speech to the instructor on the day of presentation.

Each time before class one student will broadcast news, either international, domestic, or even campus news, for about 2 or 3 minutes. Alternatively, two students can simulate a TV show, in which a host will ask a guest to comment on a piece of news or an issue under public discussion.

Course Schedule

Week	Contents
1	Orientation; PS-1 Speaking in Public; PS-2 speaking confidently and ethically; View sample speeches.

2	Chap3 Making your first speech; (Introducing a classmate or yourself)
3	PS-4 Selecting a Topic and a Purpose; PS-6 Supporting Your Ideas; OE-11 What Is Success?; mini speech
4	PS-7 Organizing the Body of the Speech; PS-8 Beginning and Ending the Speech;
5	PS-14 Speaking to Inform; PS-6 Gathering Materials; OE-6 Fake and Shoddy Goods; mini speech
6	Students' speeches: Informative Speech; 3-4 min. x 24 = 2 class hrs, plus listeners' feedback
7	PS-15 Speaking to Persuade; OE-8 Harmony Between the Young and the Old; mini speech
8	PS-16 Methods of Persuasion; OE-10 Do Appearances Really Matter?; mini speech
9	Students' Speeches: Persuasive Speech (1): 4-5 min. x 24 = 144". (Classes 1-8)
10	PS-11 Delivery; OE-7 Man and Technology (2hrs)
11	PS-10 Outlining the Speech; Speak extemporaneously
12	Students' Speeches: Persuasive Speech (2): 4-6 min. x 24 = 144 min. (Classes 1-8)
13	PS-11 Using Language; OE-14 Advertising
14	PS-17 Speaking on Special Occasions; Speaking in competitions OE-13 Man's Best Friends
15	Review and summary
16	Final exam

Course Grading

Class participation (10%)
   Attendance + Interaction (5%)
   Peer Evaluation (5%)
Speech making 60% (peer evaluation 50%+instructor evaluation 50%)
   Informative speech (20%)
   Persuasive speech 1 (20%)
   Persuasive speech 2 (20%)
Self critique and reflection (10%) (Students will view their own speeches and write critiques on them)
Final exam 20%

Useful Resources

The Art of Public Speaking online learning center: www.mhhe.com/lucas8
(to download PS study questions, flashcards, chapter objectives, outline and summary, etc)
Videos of Famous Speeches
Face to Face with Obama Shanghai, Nov. 2010,
http://www.tudou.com/programs/view/TKsYEdIsUgw/
Speeches in US history (to be available in Room 207, SEIS)
Top 100 American Speeches of the 20th Century
http://highered.mcgraw-hill.com/sites/007256296x/student_view0/top_100_speeches_.html#
TED talk: http://www.ted.com/talks

Appendix 22. The Instruction and the Evaluation Criteria for the Persuasive Speeches in Linda's Class[11]

Persuasive Speech Assignment #1

Purpose

The purpose of your first persuasive speech is to give you an opportunity to speak persuasively to the class regarding a topic you feel strongly about and will have researched and know well.
You are to apply the <u>principles of speech organization, delivery, and persuasion</u> (including argumentation).   These areas are covered in your readings and/or class lectures to date.
This assignment requires that you focus on the <u>organization</u> and <u>content</u> of your speech as well as <u>delivery</u>.
Your speech should clearly show preparation, research, organization and persuasive strategies.

Requirements

Your task is to persuade your audience to accept an opinion or take a specific action regarding an issue of your choice. You are required to follow problem-solution organization pattern/problem-cause-solution or Monroe's Motivated Sequence.   Remember that you are trying to organize a persuasive argument, so appropriate research is crucial. You will be given 4-5 minutes to complete your speech.   Practice and preparation are mandatory to ensure it will meet the time limit.

References

Include a minimum of **three** credible, published sources in your speech (research must have a broad scope--e.g., periodicals, books, journals, pamphlets and interviews--You must use at least **two** <u>different</u> types of sources)
Only **one** source may be from an internet website. It is fine to access data base information or other reference material on-line as appropriate sources.
Provide a bibliography in your preparation outline, APA format.
Verbally announce your sources OUT LOUD in your speech.   Points will be deducted if fewer than 3 citations are heard.

Persuasive Speech Evaluation Criteria

Total Points: 100

<u>Topic</u> (10 pts.)
   Relevant and appropriate
   Appealing and interesting
<u>Introduction</u> (20 pts.)
   Gained attention
   Showed relevance of topic to audience

---

[11] The instruction and the marking criteria for the two persuasive speeches were the same.

Established credibility
Introduced topic/thesis statement clearly
Previewed body of speech
Body (30 pts.)
   Structure
      Main points clear and Argument clearly developed
      Demonstrated persuasive organization
      Presented a responsible argument
      Language precise, clear, powerful
      Transitions effective
   Supporting materials
      Strong evidence presented
      Sources fully cited
      Reputable sources incorporated
      Sufficient number of sources cited
Conclusion (15 pts.)
   Audience prepared for conclusion
   Purpose and main points reviewed
   Closed speech by reference to intro./other devices
Delivery (15 pts.)
   Maintained eye contact
   Used voice, diction, & rate for maximum effect
   Used space, movement and gestures for emphasis
Overall Impression (10 pts.)
   Topic challenging
   Adapted to audience
   Maintained time limits
   Evidence of preparation & practice
 Was persuasive

Appendix 23. The Evaluation Criteria for the Informative Speech in Linda's Class

Informative Speech Evaluation Criteria

Total Points: ___/100

Introduction（4x5）   ___/20
   Gained attention
   Showed relevance of topic to audience
   Established credibility
   Introduced topic/thesis statement clearly
   Previewed body of speech
Body (5x6)   ___/30
   Main points clear
   Organization effective
   Language precise, clear and powerful
   Transitions effective
   Sufficient number of sources cited
Conclusion (6x3)   ___/20
   Audience prepared for conclusion
   Purpose and main points reviewed

Closed speech by reference to intro./other devices

Delivery (5x4)   ___/20

  Maintained eye contact
  Used voice, diction, rate & gestures for maximum effect
  Maintained time limits
  Speak extemporaneously (use note cards/outline only)

Overall Impression   ___/10

  Topic challenging/interesting
  Adapted to audience
  Evidence of preparation & practice
  Was informative
  Additional Comments (if any):

Appendix 24. Students' Peer-Marking Sheets for the Three Prepared Speeches in Linda's Class

Peer Evaluation Form: Informative Speech   Evaluator _____

Speaker		Introduction 20 pts.	Body 30 pts.	Conclusion 20 pts.	Delivery 20 pts.	Impression 10 pts.	Total 100 pts.
1							
2							
3							
4							
5							
6							
7							
…							
…							
24							
25							

Write your additional remarks, if any

Peer Evaluation Form: Persuasive Speech One[12]   Evaluator _____

Speaker	Topic 10 pts.	Introduction 20 pts.	Body 30 pts.	Conclusion 15 pts.	Delivery 15 pts.	Impression 10 pts.	Total 100 pts.
1							
2							
3							
4							
5							
6							
7							
8							
...							
...							
24							
25							

Write your additional remarks, if any.

---

[12] The same peer evaluation form was used for the second persuasive speech.

Appendix 25. The Final Exam for Linda's course

Format: Impromptu speech

Each student will be given a topic 5 minutes prior to his/her delivery of speech. S/he will have 5 minute to prepare and around 2-3 minutes to present. During the preparation time students can jot down notes on a piece of paper and they are allowed to refer to their notes when delivering the speech.

Grading Criteria

Content (30pts)
   Relevant and appropriate
   Main points adequately and logically development
   Appealing and interesting
Organization (20pts)
   Speech has a clear structure (beginning, body, and conclusion)
   Effective use of signpost words
   Transitions effective
Language (30pts)
   Speaking fluently
   Language accurate, clear, precise, and powerful
Delivery (20pts)
   Voice loud enough to hear without constraint
   Maintain eye contact
   Manipulate voice, diction, and rate effectively
   Use natural gesture